A COMPARATIVE EVALUATION OF VOICE VERSUS KEYPAD INPUT
FOR MANIPULATING ELECTRONIC TECHNICAL DATA FOR
FLIGHT LINE MAINTENANCE TECHNICIANS

THESIS

David A. Chapman, B.S.   James R. Simmons, B.S.
Captain, USAF   Captain, USAF

AFIT/GAL/LAL/95S-2

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio
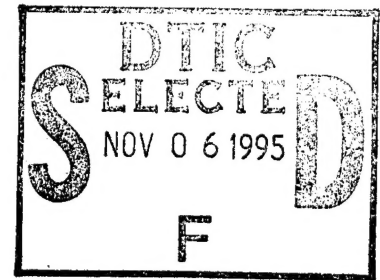
DTIC
SELECTED
NOV 0 6 1995
F

A COMPARATIVE EVALUATION OF VOICE VERSUS KEYPAD INPUT
FOR MANIPULATING ELECTRONIC TECHNICAL DATA FOR
FLIGHT LINE MAINTENANCE TECHNICIANS

THESIS

David A. Chapman, B.S.   James R. Simmons, B.S.
Captain, USAF   Captain, USAF

AFIT/GAL/LAL/95S-2

19951102 084

The views expressed in this thesis are those of the authors
and do not reflect the official policy or position of the
Department of Defense or the U.S. Government.

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

AFIT/GAL/LAL/95S-2

A COMPARATIVE EVALUATION OF VOICE VERSUS KEYPAD INPUT

FOR MANIPULATING ELECTRONIC TECHNICAL DATA FOR

FLIGHT LINE MAINTENANCE TECHNICIANS

THESIS

Presented to the Faculty of the Graduate School of Logistics
and Acquisition Management of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Logistics Management

David A. Chapman, B.S.   James R. Simmons, B.S.
Captain, USAF   Captain, USAF

September 1995

Acknowledgments

Finally, we would like to thank our wives Jessica Chapman and Mary Jo Simmons. These two ladies were the model of support and for that we thank you. A special thanks also goes to Dave's girls, Kendall and Courtney, and to Jim's boys, Andy, Tyler, and Jordan. We are both looking forward to having true family lives again soon.

<div align="right">

Dave Chapman

Jim Simmons

</div>

Table of Contents

List of Figures

List of Tables

AFIT/GAL/LAL/95S-2

Abstract

Interactive Electronic Technical Manuals will soon become a requirement for aircraft maintenance technicians. An important aspect in their development is the selection of an input device that will enhance, rather than impede, technician performance. The purpose of this thesis was to evaluate two types of input devices that can be used: a voice recognition input and a keypad input. Studies to date have evaluated the superiority of digital data over paper data, and advantages of using a Head Mounted Display Device over a flat screen laptop computer. No research has evaluated the input device. An experiment was conducted to determine which interface allowed the technicians to work faster. Sixteen F-16 avionics maintenance technicians from the 178th Tactical Fighter Group, Ohio Air National Guard, performed two parallel tasks using each input device. One task was performed using a keypad input device and another task was performed using a voice recognition input device. Raw data showed no statistical difference in task completion times between input devices. However, when computer processing time was subtracted from the voice task times, there was a slight time difference found. Most importantly, results indicate that the technicians liked the advantages of the voice recognition input device over the keypad input device. The primary conclusion is that voice recognition may be a desirable input configuration and further study is warranted in more stringent environmental conditions.

A COMPARATIVE EVALUATION OF VOICE VERSUS KEYPAD INPUT FOR
MANIPULATING ELECTRONIC TECHNICAL DATA FOR FLIGHT LINE
MAINTENANCE TECHNICIANS

I. Introduction

Chapter Overview

The modern weapon systems in use today require much more technical data than in the past. For example, the B-1B bomber has over one million pages of technical data. This, coupled with the decrease of personnel in the aircraft maintenance career field, makes it imperative that a more efficient means of displaying and manipulating technical data be developed. This chapter discusses the general issue of the need for an efficient means of displaying and manipulating technical data, followed by the specific problem statement we intend to follow in this thesis. From this problem statement, we outline our research objective and hypotheses. This chapter concludes with a discussion of the scope and limitations of our study.

General Issue

Aircraft maintenance technicians rely on a technical order (TO) for every task performed in maintaining an aircraft. It has been recognized for many years that conventional technical orders used to support maintenance personnel are incomplete, poorly organized, and difficult to use. With the increasing complexity of aircraft, more and more technical orders are required to maintain the aircraft. The technicians are inundated with a sea of paper instructions that are cumbersome not only to use but also to get to the job site. Many problems are encountered with the accuracy of the TOs, as times are often long before corrections reach the field. With a paper TO, it is tempting for the

technician to set the TO aside and instead rely on personal experience. Automation of the technical order system has appeared to be the logical solution to these problems (Thomas and Clay, 1988).

The Air Force and the Department of Defense have been moving toward digital data since the early 1980s. Interactive electronic technical manuals have been developed that allow aircraft maintenance technicians to view needed technical data on a portable computer that can be held in their hands. Current research is underway to link all maintenance information systems together under one standard human-computer interface. The modern aircraft technician will not only have to master the increasing complexity of new weapon systems, but also the numerous information systems that go with them. Making technical data more available and easier to use will help technicians keep ahead of this challenge.

In a 1988 memo, the Deputy Secretary of Defense directed the military departments and the Defense Logistics Agency to employ Continuous Acquisition and Life-Cycle Support (CALS) technology for all new weapon systems and, where feasible, for weapon systems currently under development (Clark, et al., 1992). The objective of CALS is to improve the productivity and quality in acquisition and logistics support of DoD weapon systems thereby improving readiness and operational effectiveness and reducing system life cycle costs (Department of the Air Force, 1993). This re-emphasis of CALS technology brings current logistics research to the forefront. When the F-22 is fielded, along with it will come an integrated maintenance information system linking all facets of maintenance together, beginning with the automated presentation of technical data. Ensuring the technicians have an effective means of displaying and interacting with this technical information is vital to maintaining the aircraft. This change in the presentation of data will have a significant impact on the way aircraft maintenance is performed.

Specific Problem

The technology specified in the CALS directives for electronic technical manuals currently exists. Armstrong Laboratory has developed a Head Mounted Display Device (HMDD) capable of displaying digitized technical data. The current configuration consists of a lightweight vest with a small drive for storage of data, two 12 volt batteries for the power supply, and a keypad type input device. The display device used is a miniature VGA display capable of displaying both text and graphics, and projecting an image equivalent to that of a 12 inch computer at two feet.

Research performed on the current configuration has suggested that there are limitations to the effectiveness of its use. The interface is a keypad type that requires the technician to use specific buttons to manipulate or move through the technical data. Use of a keypad for input requires a shift in focus from the display device to the keyboard back to the display, forcing the technician to take his focus away from the technical data and the task at hand. The keyboard also requires that the technician take a hand away from the task to input commands to the device. This shift of focus and requirement for use of the hands is often impractical, if not nearly impossible, when performing maintenance on the flight line. As a possible solution to this problem, Armstrong Laboratory has suggested the use of voice recognition as a means of input to the device to alleviate these limitations and possibly improve technician performance. Machines that occupy the operator's hands and eyes become more efficient with voice technology (Poock, 1980). These electronic interfaces are more efficient than the keyboards and push buttons normally used to control machines (Berardinis, 1993). Specifically, Armstrong Laboratory is interested in determining if the addition of voice recognition technology to the current HMDD will enhance flight line technician performance.

Research Objective

The objective of this thesis research is to determine the extent and nature of any performance differences between technicians accomplishing maintenance tasks using keypad versus voice input to manipulate digitized technical data presented on a HMDD.

Experimental Hypothesis

The overall research hypothesis is that technician performance will be enhanced by using voice recognition as an input when compared to keyboard entry as an input for technical data displayed on a HMDD. The following hypotheses further refine the overall research hypothesis and serve as the basis on which to compare technician performance:

1. Task completion times using voice recognition will be faster than task completion times using the keypad.

2. System performance with the voice input configuration will meet accepted industry standards.

3. User satisfaction will be greater with the voice input configuration than with the keypad input configuration.

Scope and Limitations

The hardware and software used in this experiment is limited to that currently used by Armstrong Laboratory. The HMDD is the current display device being used by Armstrong Laboratory. The current configuration consists of the monocular display attached to a standard crew chief protective helmet and a small microphone, plus a lightweight vest weighing approximately 10 pounds that holds the battery pack, computer memory and CPU, and the keypad. The software used for the addition of the voice recognition capability is VoiceAssist by Creative Labs, Inc. This is a commercial off-the-shelf product selected by the engineers at Armstrong Laboratory. It is a speaker

dependent software system. No alternative designs of the HMDD were considered, and no other available voice recognition software packages were evaluated for use.

The tasks for this research will be performed at the Springfield, Ohio Air National Guard (OANG) unit by F-16 aircraft maintenance technicians. All technical data currently in digitized form is for the F-16 aircraft. The Springfield Guard unit is the closest F-16 unit. The tasks are limited to two flight line maintenance tasks performed by flight line avionics maintenance technicians. The flight line maintenance environment provided the maintenance environment most challenging for the HMDD and the voice recognition capability. The tasks were limited further to one of three aircraft subsystems of the F-16 for which digitized technical data had already been authored: the Inertial Navigation System (INS), the Fire Control Radar (FCR), and the Heads-Up Display (HUD). The length of the task will be limited by the battery life of the HMDD. The task will be limited to approximately 30 minutes to ensure that technicians will not have to stop in the middle of a task to replace the battery.

II. Background

Chapter Overview

As the size of the military is decreased to meet end strength force requirements, certain steps must be taken if we are to maintain the current level of capability. To draw from the common saying, we will be forced to do more with less. To help the service accomplish this task, technology can be applied to certain Air Force applications. For example, the development of the Integrated Maintenance Information System will help aircraft maintenance technicians work more effectively and efficiently. This system combines several existing maintenance databases. The system provides technical information, historical information, and ties into the base level supply system. Taking full advantage of technology will provide the greatest benefit by not necessarily allowing fewer people to accomplish more work, but will allow each individual to be more productive. In an effort to make the performance of tasks more efficient, the portable maintenance aid (PMA) was developed. This is a very effective tool but one important aspect, the input device, has been neglected during its development. The objective of this thesis is to evaluate two different input devices to determine if technician performance using the PMA can be improved by adding speech recognition to the current configuration.

This chapter is divided into three main sections. The first section, system development, traces the incremental steps taken in the development of the portable maintenance aid. The second section focuses on our assertion that voice recognition should be added. Research leading to the development of the multiple resource theory is examined, supporting the idea that performance can be improved by using multiple input channels to perform a task. Following this explanation, research comparing user performance while using voice recognition will be addressed.

System Development

The Air Force Human Resources Laboratory (now Armstrong Laboratory) has been conducting research and development for an automated technical data presentation system since 1976. A summary of this research is shown in Table 2-1. This research was initiated because of potential performance improvements and the potential reductions in the cost of maintaining the Air Force Technical Data System (Thomas and Clay, 1988). Two preliminary design studies were performed by Armstrong Laboratory (AL) in the late 1970s to determine the feasibility of an automated presentation system for aircraft maintenance technical data. The results provided information for the development of a prototype presentation system that could be used in a field demonstration of an intermediate level prototype.

Throughout the development of the prototype system, emphasis was placed on three areas. In the early development of the system, one primary concern was the presentation of the data in electronic form. It was very difficult to present schematics and wiring diagrams in an acceptable format. The second area of concern was the user acceptance of the system. In the later development stages, the emphasis was on the type of display that could be used and how it would improve the overall usability of the system. It was not until 1993 that any formal research was done on the user interface, which only evaluated the usefulness of the existing graphical user interface (Carney and Quinto, 1993).

The early development of a prototype presentation system began in 1982 with the Computer-based Maintenance Aids System I (CMAS I). This system was followed by CMAS II in 1985. These two projects focused on developing human factors and data presentation requirements (Thomas and Clay, 1988). In the CMAS I project, a MODCOMP Model 7840 minicomputer with a standard keyboard interface was installed

Table 2-1. Development of the Portable Maintenance Aid

Study	Configuration	Results	Recommendations
CMAS I, 1984	MODCOMP Model 7840 minicomputer	1. Did not gain user acceptance 2. Extremely slow response time	1. Decrease the response time of the system
CMAS II, 1985	Grid Compass Model 1139 microcomputer with standard keyboard	1. Response time good 2. Technicians could effectively use system	1. Use larger display 2. Improve schematics 3. Make more portable
PCMAS, 1989	Semi-ruggedized portable computer with standard keyboard	1. Technicians successfully used system 2. Technicians thought system provided easy access to related data	1. Build portable system small enough to use in areas of aircraft inaccessible to PCMAS
Masquelier, 1991	HMDD connected to desktop computer compared to computer with flat panel display with standard keyboard	1. No statistically significant performance differences between display devices.	1. Evaluate HMDD on flight line maintenance tasks
Friend and Grinstead, 1992	Fully portable HMDD compared to hand-held portable computer, both using dedicated hardware keys, push button keys, cursor keys, and number keys	1. Tasks completed faster with HMDD in cockpit task 2. More faults detected with HMDD in engine task	1. Test on more complex maintenance tasks 2. Test on more complex weapon system
Carney and Quinto, 1993	Personal laptop computer with programmable soft-keys, dedicated hardware keys, push button keys, cursor keys, and number keys	1. Dedicated hardware keys and number provided greatest user satisfaction 2. Pushbuttons and programmable soft keys provided lowest user satisfaction	1. Test different types of input devices, such as mouse or touchscreen 2. Evaluate same interface in a different working environment, such as flight line

in an intermediate level avionics maintenance shop to collect performance data and user opinions. In August 1985, the CMAS II prototype, consisting of a GRID Compass microcomputer with a standard keyboard for input, was placed in an intermediate level

shop for a 2-week field demonstration. The objective was to get preliminary indications of the effectiveness as compared to the paper documentation and to determine the overall acceptance of the system (Thomas and Clay, 1988).

In the later development of the system, the focus shifted from data presentation to refining and testing the unit in realistic conditions. Three studies were performed evaluating the unit in a realistic environment with a goal of establishing a useful unit for flight line aircraft maintenance. The first study used a small, rugged, portable computer with a flat panel display and standard keyboard input to evaluate performance and usability with flight line maintenance technicians (Thomas and Clay, 1988). The second study compared the flat panel display to a monocular Head-Mounted Display Device (HMDD) for technicians working in a support shop environment (Masquelier, 1991). Performance of the technicians with each device was measured. The third study was similar to the Masquelier study, but performed the experiment in a flight line environment (Friend and Grinstead, 1992). The last two studies used a keyboard type input, but the standard keyboard was changed to include a mixture of dedicated hardware keys, cursor keys, number keys, push-button keys, and programmable soft-keys.

In 1993, a study evaluating the usefulness of the interface used for the HMDD was performed. This study focused on which of the existing features of the interface enable users to access the information with the highest degree of satisfaction. (Carney and Quinto, 1993). The researchers hoped that by identifying the best features, redundant features could be eliminated. The study was able to identify the best features of the existing interface. However, alternative interface designs were not examined.

Through many studies, a portable display device was developed and proven feasible and effective for displaying digitized aircraft maintenance technical data. Much effort was expended in making the unit applicable to the flight line maintenance activity, resulting in a unit that can be used in the small, inaccessible areas often encountered in

maintaining aircraft. Throughout the development of the system, one important aspect, the input device, was neglected in the push to improve technician performance. Many studies have shown that interfaces appropriate for the environment and tasks can significantly affect performance. These studies are discussed later in this chapter. Improvements in technology have made options available that were not possible even a few years ago. For example, voice recognition technology has developed to the point where it has become an acceptable computer interface method. Voice recognition allows a user to command a machine without the use of hands. This concept shows many potential benefits. For example, in a maintenance environment, tasks often require the use of both hands. A computer interface which allows the technician free use of both hands has the potential to greatly improve task performance. AL has suggested using a voice recognition interface with the HMDD to free the technician's hands during maintenance. The following paragraphs review the feasibility of using voice recognition as an input means and the supporting research.

Input Device

This paper proposes that using voice recognition will improve performance when compared to the existing keypad device. This research will examine the input device by adding voice recognition capability to the current configuration of the HMDD. Technician performance using voice will be compared to performance using the keypad. Our proposal was suggested by AL and is supported by research in the literature. This section reviews the use of voice recognition as a viable input channel. Research supporting the Multiple Resource Theory and human performance improvement will be cited as justification for this study.

Use as an additional input channel

The viability of voice recognition has been reported in numerous studies pertaining to its usefulness as an input channel. The military as well as the civilian sector have reported a great deal of interest in this new technology potential. In 1992, as a possible lead in to other applications, an experiment was performed adding speech recognition to InterFIS, a natural language interface to the troubleshooting module of the fault isolation shell (FIS). FIS is an expert system development tool for the diagnosis of failures in analog electronic equipment. FIS computes the probability that a particular fault hypothesis is correct after a test has been performed, and then recommends the next best test based on the information supplied by the technician. The original interface was a combination of keyboard input and graphic displays. The addition of speech recognition capabilities was found to significantly enhance the friendliness and ease of use. Researchers found that subjects prefer spoken to typed input because spoken input is faster. This study did not attempt to measure the change in productivity caused by a change in input. The study only tested to see if the addition of the speech recognition to the interface was successful (Everett, 1992).

Numerous other studies have been performed focusing along similar lines. A summary of the more notable studies can be found in Table 2-2. When examined as a whole, the conclusion is quite clear: voice recognition has come of age and is a viable input and manipulation method for human machine interface. The application of voice recognition technology to human machine interface is not an arbitrary occurrence. The theoretical roots of voice recognition can be found in the Multiple Resource Theory.

The Multiple Resource Theory

The theory behind the benefits gained by voice recognition was coined by Christopher Wickens in 1981. The theory is known as the Multiple Resource Theory, and it supports the concept that individuals can speak and work at the same time. Two studies

Table 2-2. Summary of Additional Input Channel Studies

Study	Purpose	Results
Reed, 1982	Introduction of voice recognition into Army helicopters	Voice recognition can be used in high noise levels given special attention in training the system
Martin, 1989	Compared performance of speech input with typed input and mouse clicks	Speech input more efficient response channel for input and manipulation of data
Schmandt, Ackerman, and Hindus, 1990	Examined usefulness of speech to control window navigation in a Windows-based system	Speech found superior to mouse when windows were partially or completely obscured
Pausch and Leatherby, 1991	Evaluation of the utility of speech input to graphical editors	Speech benefit found when using voice input in parallel with mouse
Everett, 1992	Evaluation of speech recognition added to InterFIS	Speech recognition improved the ease of use and was the preferred method of entry by subjects
Karl, Pettey, and Schniederman, 1993	Evaluated advantages of using speech recognition over mouse for word processing applications	Performance times 18.7% faster using speech input over mouse input
Manaris, 1994	Viability of speech recognition shown by developing a natural language interface for the UNIX operating system	User more productive using a natural language interface

by Wickens and Wickens et al. finalized the development and refinement of the Multiple

Resource Theory. In these two studies, people are asked to perform two tasks

simultaneously, such as tracking a target and entering data into a computer. It is shown

by the researchers that the separate tasks tend to interfere with each other; but this

interference is minimized when the tasks are spread across multiple modalities, or mental

resources, such as speech and typing. Both studies recommend that when computer

interfaces are built for multiple simultaneous tasks, speech input capabilities may be

effective in enhancing user's abilities to perform the multiple tasks efficiently (Wickens,

2-7

1980 and Wickens et al., 1981). Wickens synthesizes his and previous research in the field into the multiple resource theory of attenuation:

> The brain modularizes the processing of different types of information. When different tasks tap different resources, as manual movements and speech are thought to do, then much of the processing can go in parallel, not interfering with the other. When the tasks tap the same resource, interference between the tasks occurs and processing slows. (Wickens et al., 1981)

This theory helped justify further research into voice recognition as an additional input channel for computer operation.

Several studies have been done which validate the idea that using multiple input channels can be better than using only a single input channel. These studies are summarized in Table 2-3. Having established a theoretical base, the next logical step is to focus on applications of the theory. The Multiple Resource Theory says that we can process data through the brain in parallel. This fact can be exploited in various situations to improve human performance.

Performance improvement

User performance, defined as the time required to accomplish a task, improves using voice recognition. This concept is supported in the literature by numerous studies that show performance improvements using voice recognition. One of the earliest and most important user performance studies was accomplished by Poock in 1980. Poock compared the speed and accuracy of speech and typed entry of command and control inputs. Twenty-four military officers were observed logging into several different host computers, reading messages, deleting files, and transferring files, using both typed input as well as speech input. Speech input was found to be 17% faster than typing, and typing produced 183% more errors than speech. The study also showed that the users preferred speech input over typed input (Poock, 1980).

Table 2-3. Development of the Multiple Resource Theory

Study	Purpose	Results
Allport, Antonis, and Reynolds, 1972	Auditory shadowing – subjects hear and repeat word while studying pictures for a memory test	Subjects could focus on two tasks at the same time
Treisman and Davies, 1973	Two pair of stimuli simultaneously presented Memory and target detection studied	Presentation split between eyes and ears. Performance higher than with eyes or ears alone
Sternberg, et al., 1978	Looks at typists' ability to use both hands	Experienced typists are faster at typing words involving the use of both hands rather than only 1
Larochelle, 1984	Typing study	Same results as Sternberg, et al, 1978
Wickens, 1980 and Wickens, et al., 1981	Simultaneous tasks – tracking a target and entering digits into a computer	Dual task interference is minimized when the tasks are spread across multiple modalities

Not every aspect of speech recognition is always found to be positive or beneficial, but generally some type of performance improvement is found in every study. A summary of other performance studies can be found in Table 2-4.

As Table 2-4 shows, there are many performance benefits to be realized by the implementation of high quality, reliable speech recognition systems. "Speech input could potentially provide effective dual task performance improvements with a typical direct manipulation task" (Quill, 1993). The results of previous research do not always support the claim that speech input is faster than any other mode. However, the overall results suggest that speech input can provide a valuable additional response channel in situations where a user's hands are likely to be busy performing another task.

Table 2-4. Summary of User Performance Studies

Study	Purpose	Results
Poock, 1980	Compared speed and accuracy of speech vs. typed entry of command and control computer inputs	Speech was found to be 17% faster and typing produced 183% more errors
Elster, 1980	Examined the effects of background noise on user performance with a speech recognition system	Background noise could be adjusted for when training the system for use
Cochran, Riley, and Stewart, 1980	Complex technical information entered using keyboard and using a speech input device	Speech input took longer but produced fewer errors
Nye, 1982	Compared speech to keyboard entry of destinations in an airplane baggage sorting task	Keyboard entry had a higher error rate
Poock and Martin, 1984	Examined effects of operator stress on the accuracy of a voice recognition system	System training can drive errors to a nominal level
Leggett and Williams, 1984	Compared programming performance using speech and keyboard input devices	Keyboard entry was faster but had a higher error rate
Visick, Johnson, and Long, 1984	Compared speech and keyboard input devices for entering destinations in a parcel sorting task	When hands were busy speech yielded a 37% improvement in time but had a 40-80% error rate

Conclusion

Research to this point supports the notion of improved performance with voice recognition, but thus far no study has evaluated this voice recognition capability in a real world, "hands busy" environment where mobility is also required. The viability of voice recognition as an input alternative has been established in the literature, as has the possibility of performance improvement while using voice recognition. The research objective of this thesis is to tie the performance advantages available through speech recognition into the performance advantages available from using digitized technical data for flight line maintenance. Our hypothesis is that technician performance will be better

when using speech recognition than when using a keypad input channel. Based on previous research in the field of human performance, we expect to find this to be true. If a performance improvement is indeed found, a cost benefit analysis should be performed examining the feasibility of implementing this capability into the next generation of maintenance aids. The next chapter will explain the methodology used for the performance of this experiment, including the experimental design and possible limitations of the study. Attention will also be given to the statistical techniques used to analyze the performance differences found.

III. Methodology

Chapter Overview

Portable Maintenance Aids (PMAs) can be excellent tools to aid the aircraft maintenance technician. Unfortunately, the development and testing of these PMAs has not examined the advantages and disadvantages of alternative input devices such as a keypad, voice recognition, or a mouse. The goal of our study is to determine the effect of an alternative input source on technician performance. This chapter will explain the experimental methodology necessary to evaluate the differences in technician performance. First we discuss the experimental design and the hypotheses we are testing. Next we describe the equipment used in the experiment. Then we discuss the tasks and experimental subjects chosen for examination. Next we explain the data collection and analyses necessary to support or refute our hypotheses.

Experimental Design

A total of 16 maintenance technicians performed two different troubleshooting tasks on the Heads Up Display (HUD) system of the F-16C/D. Both tasks were performed with technical data displayed on a HMDD. One task was performed using a keypad input device to manipulate the data and the other task was performed using a voice input device to manipulate the data. Following is a discussion of the experimental variables and controls as well as a discussion of the experimental design used.

Variables

This experiment examined a single independent variable. Previous research has examined the effects of presentation media, display device, and the experience level of the technicians. Because these effects have already been examined, they are not of interest in this research effort. The independent variable of interest in this study was input device.

3-1

The two levels of this independent variable were the keypad input device and the voice input device. The dependent variables used to determine the effect of the independent variable were task completion times and command input errors. Task completion time was measured from the first command to the completion of the maintenance task. Command input errors were defined as any command given to the computer that was not properly recognized or executed.

Controls

An experimental plan, shown in Appendix A, was developed for the experimenters to follow during test sessions. The plan was used to standardize the presentation of instructions and troubleshooting problems for all test subjects. The same experimenters conducted all of the data collection activities. All data collection runs were performed at the same location. All subjects were randomly assigned to one of two experimental groups. All subjects were tested for 20/20 corrected or uncorrected vision. Subjects were asked to perform a test to determine their dominant eye. Each test subject received identical training on how to use the system, as well as how to use each input device. Computer response times for the voice recognition software were determined and subtracted from the voice task completion times. System errors, including both maintenance and computer errors, were all handled identically. If an error was made, the task was halted and re-started at the same place the error occurred. Additionally, learning effect was controlled for by alternating which input device was used first by each test subject. Half of the test subjects used the voice input first and the other half used the keypad input first. Also, order effect was controlled for by alternating which maintenance task was performed first by each test subject. Half of the subjects completed Task One first and the other half completed Task Two first. A pilot study was performed using two maintenance technicians to evaluate the sequence of events in the experiment, to validate

the troubleshooting steps, and to determine the amount of time required to complete the experiment.

Latin Square

The Latin Square Design was selected for use in this experiment because it allowed us to determine the main effect of the data manipulation device on technician performance (Neter, et al., 1985). The 16 subjects were randomly divided into two groups. Each member of each group performed both tasks. One task was performed using voice input to manipulate tech data and the other task was performed using keypad input. The task performed using voice input was alternated between groups. This design is shown in Table 3-1.

Table 3-1. Latin Squares Design

	Voice Input	Keyboard Input
Group 1	Perform task 1 first	Perform task 2 second
Group 2	Perform task 2 first	Perform task 1 second

Experimental Hypotheses

The following hypotheses served as the framework for comparing technician performance using the two input devices.

Hypothesis I

This hypothesis predicted that the task completion times using the voice input device would be faster than the task completion times using the keypad input device.

Hypothesis II

This hypothesis predicted that the system performance with the voice input device would meet accepted industry standards.

Hypothesis III

This hypothesis predicted that user satisfaction would be greater with the voice input device than with the keypad input device.

Hardware

The computer used in this experiment is a 486-based platform operating at 33 MHz with 16 MB of RAM. It is a complete single board computer with a 180 MB removable hard drive. It supports VGA graphics with a video resolution of 640 x 480 pixels. The computer also contains a Sound Blaster audio card with a stereo digitized voice channel. The eye-piece used in this experiment is a HMDD manufactured by Imaging and Sensing Technology. It supports a 640 x 480 pixel monochrome display. This display projects an image of a 12 inch computer screen viewed at 2 feet. The eye-piece is mounted to a standard crew chief protective helmet. See Figure 3-1 below.

Figure 3-1. Equipment Configuration

The microphone used in this experiment can be used with or without a noise filtering device. When used without the filtering device, the microphone is a simple dynamic

microphone. When used with the noise filter, two microphones are used with a differential summing amplifier to cancel out common mode noise. The filter is a 1500 Hz low-pass filter with a 6 dB per octave cutoff.

Software

The software system used for presentation of the digitized technical data, PCIMIS, was developed jointly by the Armstrong Laboratory and CSERIAC, the Crew System Ergonomics Information Analysis Center. CSERIAC is a DoD information analysis center operated by the University of Dayton Research Institute. PCIMIS is a Windows-based application designed specifically for the purpose of displaying technical data and interfacing with other maintenance data collection systems, such as CAMS, REMIS, and SBSS. This presentation system displays both text and graphics, including the large schematics and wiring diagrams found in paper TOs. The voice recognition software used in this experiment is called VoiceAssist. It was developed by Creative Labs Inc. in Milpitas CA. It is a commercial, off-the-shelf package that can be used on almost any personal computer. It is a speaker dependent software system. Each user was required to train the software to recognize his/her voice before the system could be used to perform the task. VoiceAssist allows users to navigate the Windows environment and run Windows applications using voice commands. It supports multiple users, each having their own command set (Davenport, 1993).

Tasks

Two aircraft maintenance tasks were required for the evaluation of the experimental hypotheses. Several criteria established by the researchers for selecting appropriate maintenance tasks are discussed, followed by a discussion of the actual tasks chosen.

3-5

Criteria

There were six criteria identified for considered for task selection. First, the two tasks should be parallel. Parallel tasks were defined as two tasks that were equal in difficulty and required the same skills to complete. Next, they should by representative of routine maintenance performed, meaning that the the tasks are normally encountered in the performance of daily maintenance. The tasks should be an appropriate length. The task lengths were required to be less than 30 minutes in order to not exceed the battery limitations of the system. Tasks should require the use of both hands and should require the technician to move about the aircraft. This is required to effectively evaluate the performance differences found between the two input devices. Finally, it was desirable that tasks be chosen for which presentation data was already developed.

Selection

Three systems of the F-16C/D Fighting Falcon were available for evaluation: the Inertial Navigation System (INS), the Heads-Up Display (HUD), and the Fire Control Radar (FCR). Technical data for these three systems was already in the format required for the presentation system. Two HUD tasks were chosen for this experiment. These two tasks were deemed parallel and routine during previous research conducted for the user field test and demonstration conducted at Luke AFB, AZ during the summer of 1994 (Thomas, 1995). Conversation with avionics system experts revealed that the tasks chosen would be of an appropriate length, require the use of both hands, and require movement about the aircraft. Based on evaluation of selection criteria, tasks for HUD MFL 001 and 002 were chosen for this experiment. The digital data used during the experiment is identical to paper tech data in the number and order of steps performed. However, the presentation of information is different than would be found in a paper TO. A sample screen of information as viewed by test subjects is shown below in Figure 3-2.

```
┌─────────────────────────────────────────────────────────┐
│           CHECK FWD DISPLAY MUX HUD DIGITAL DATA CKT2     │
├─────────────────────────────────────────────────────────┤
│ ❶  1.  Verify 2.0 ohms from 9472P1 Pin BB to 9472P1 Pin CC.│
│                                                           │
│          ○ OK                                             │
│          ○ NOT OK                                         │
│                                                           │
│         ┌─────────────────── Note ───────────────────┐   │
│         │                                             │   │
│         │  ⓘ   All resistance checks shall be made    │   │
│         │      with digital meter.                    │   │
│         │                                             │   │
│         │             ┌────────┐                      │   │
│         │             │   OK   │                      │   │
│         │             └────────┘                      │   │
│         └─────────────────────────────────────────────┘   │
│                                                           │
├─────────────────────────────────────────────────────────┤
│ Highlight appropriate choice, and press F1_Next to continue.│
└─────────────────────────────────────────────────────────┘
```

Figure 3-2. Sample Screen of Information

Subjects

The HUD is a sub-system of the aircraft avionics system and is the responsibility of
avionics systems maintenance technicians. The population of interest in this experiment is
all F-16 avionics maintenance technicians. In order to generalize the results of this study,
the sample chosen had to be representative of the population. To be considered
representative of the population, the sample had to routinely perform flight line
maintenance on the F-16. They also had to come from an operational unit.

The 178th Tactical Fighter Group (TFG), Springfield Ohio Air National Guard
flies the F-16C/D and was chosen to support this experiment. The 178 TFG is basically
structured like an active duty operational flying squadron with extra maintenance support
elements. Flight line and maintenance support elements fall under the same chain of
command. Active duty units are organized with flight line maintenance working for the

Operations Group and the maintenance support elements working for the Logistics Group. The number of maintenance personnel assigned, the number of assigned aircraft, and aircraft utilization rates are proportionately similar to active duty units.

The 178th has approximately 20 technicians qualified to perform maintenance on the HUD system. Maintenance technicians are required to perform their duties on a daily basis to support the flying schedule. Background information was collected on each qualified technician. See Appendix B for a sample of this collection form. Examination of the training data showed that there was no difference in training requirements between guardsmen and active duty avionics maintenance technicians. Based on the above information it was determined that the sample was representative of the population of maintenance technicians.

The 16 test subjects performed two parallel tasks. This yielded 32 data points which were ultimately broken down into two samples of 12. One sample contained the task completion times for all tasks completed using the voice input device and the other sample contained the task completion times for all tasks completed using the keypad input device. Similar studies have been accomplished evaluating this hardware and significant results were found with the same sample size.

Data Collection

Both quantitative and qualitative data were collected for the evaluation of the experimental hypotheses. Quantitative data were collected for the evaluation of Hypotheses I and II. Qualitative data were collected for the evaluation of Hypothesis III.

Quantitative

Quantitative data were collected for the evaluation of Hypotheses I and II. Task completion times were recorded to evaluate Hypothesis I. Task completion times were recorded using a timing routine built into the computer presentation system. These times

were verified by the experimenter using a stopwatch. The timing routine started a clock when the first command was given to enter the task. The clock was stopped when the technician cleared the last screen of the task. System errors were recorded to evaluate Hypothesis II. Computer errors were defined as any command given to the computer that was not properly recognized or executed. The experimenter closely monitored the commands input by the test subjects and annotated all command input errors.

Qualitative

Qualitative data were collected for the evaluation of Hypothesis III. The qualitative data consists of subjective answers to questions used to analyze user likes and dislikes of the input devices. See Appendices C, D, and E for samples of the qualitative data collection forms.

Data Analysis

Both quantitative and qualitative data were collected for the evaluation of the experimental hypotheses. Quantitative data analyses were conducted for the evaluation of Hypotheses I and II. Qualitative data analyses were conducted for the evaluation of Hypothesis III.

Hypothesis I

The quantitative data analysis required to evaluate Hypothesis I was a paired t-test. This test allowed us to compare the means of the task completion times for the voice and keypad tasks. The only assumption necessary in using the paired t-test is that the data be normally distributed. To verify the assumption of normality, the task completion times for each input device were analyzed using a Wilk-Shapiro test for Normality. Task completion times were input into Statistix, a statistical software program (Statistix, 1992). The test statistic returned was then compared to the minimum value for a 0.01 significance level with a sample size of twelve. The minimum value is 0.805 (Conover, 1980). Any

data that returns a test statistic greater than the minimum value meets the assumption of normality. Once the assumption of normality was verified, a paired t-test was performed to examine the difference in the two mean task completion times. Microsoft Excel version 5.0 was used to perform the paired t-test (Microsoft, 1994).

Hypothesis II

The quantitative data collected to evaluate Hypothesis II included system errors converted into a percentage of commands recognized properly by the computer. The number of commands accepted was divided by the number of commands given to obtain this percentage. Mean system performance was calculated by averaging the system performance rates of each test subject. This mean performance value was compared to the suggested standard of a 95 percent recognition rate (Poock and Roland, 1982).

Hypothesis III

The qualitative data collected to evaluate Hypothesis III included questions with numbered responses and open-ended questions. These questions were used to analyze the user likes and dislikes of each input device as well as the overall system. A mean was calculated for each question requiring a numbered response. These means were compared to the middle value of 5 which indicated no feeling one way or the other. Responses were divided into four categories. Questions with a mean response of between 5.0 and 6.990 were labeled as barely positive, questions with a mean response of 7.0 to 7.49 were labeled as moderately positive, questions with a mean response of 7.5 to 7.99 were labeled as positive, and questions with a mean response greater than 8.0 were labeled as very positive. Test subject responses to open-ended questions were used to gain additional insight into user preference.

Summary

The goal of this study was to determine the effect of an alternative input source on technician performance. This chapter explained the experimental methodology necessary to evaluate the differences in technician performance. The experimental design and the hypotheses tested were discussed. Next we described the equipment used in the experiment. Then we discussed the tasks and experimental subjects chosen for examination. Next we explained the data collection and analyses necessary to support or refute our hypotheses. Results and analyses of data follow in Chapter IV.

IV. Results and Analysis

Chapter Overview

This experiment was designed to collect data in three categories: task completion times, system performance results, and user responses to input devices and overall system. Sixteen maintenance technicians were randomly selected to each perform a task using a voice input device and a keypad input device, for a total of thirty two tasks. Task times were recorded for each voice task and each keypad task. Times were recorded using a timing routine installed on the computer system and then verified using times recorded by the observer using a stopwatch. In addition to task completion times, the number of commands not recognized by the computer were recorded for each voice and keypad task. Qualitative evaluations were performed by each test subject after each task performed. Each subject answered 10 questions evaluating each input device. Questions were developed using a nine point Likert scale. Open-ended questions were also answered on each input device and the overall system.

Quantitative Results

Task completion time data for each input device was first analyzed using a Wilk-Shapiro normality test to ensure that the data conforms to a normal distribution. Having verified the necessary requirement, a paired two sample t-test for means was performed on the data. System performance data was collected and converted to a percentage of commands correctly recognized by the computer. These percentages were compared against the voice recognition rate of 95% suggested by Poock and Roland (Poock and Roland, 1982). Descriptive statistics are used to evaluate qualitative data. The means of each answer are compared to the middle value (five), representative of no feeling one way or the other, to assess user perceptions of the input device.

Hypothesis I - Task Completion Times

This hypothesis predicted that task completion times using voice recognition would be faster than task completion times using the keypad. The data collected and analyzed for this portion were the overall task completion times for the voice tasks and the keypad tasks. Of the sixteen subjects tested, the last three subjects (fourteen through sixteen) did not successfully complete the tasks. There were technical difficulties with the system hardware that prevented the collection of valid test results. As a result of these problems, test results for subjects thirteen through sixteen were not used in the data analysis. The data for subject thirteen was not used because it is paired with the data for subject fourteen. Data collected for subjects one through twelve is shown below in Table 4-1.

Table 4-1. Task Completion Times

Subject Number	Voice Task Duration	Keypad Task Duration
1	7.67	21.52
2	12.08	11.62
3	10.02	12.92
4	12.5	8.42
5	7.28	11.42
6	10.85	8
7	8.73	10.42
8	10.83	8.07
9	7.52	12.82
10	12.05	7.33
11	8.63	11.7
12	12.32	6.12
Mean	10.04	10.86
Standard Deviation	1.92	3.88

A Wilk-Shapiro test for normality was performed on the data shown in Table 4-1. The data for voice task completion times returned a Wilk-Shapiro value of 0.9168. The Rankit Plot is shown in figure 4-1. Using the minimum value of 0.805, we accept the

4-2

assumption that this data conforms to a normal distribution (Conover, 1980). The data for the keypad task completion times returned a Wilk-Shapiro value of 0.8216. The Rankit plot is shown in figure 4-2. This data conforms to a normal distribution. However, further examination of the keypad task completion time data revealed the existence of a statistical outlier.

Wilk-Shapiro / Rankit Plot of VORG

Approximate Wilk-Shapiro 0.9168   12 cases
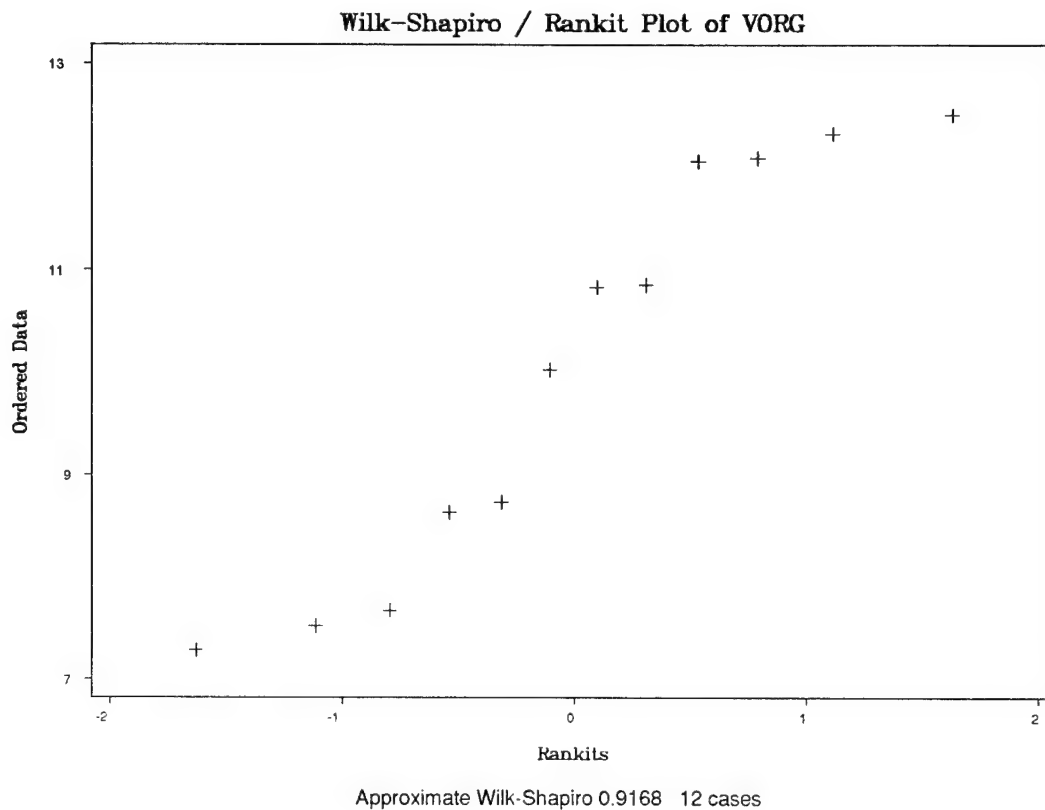
Figure 4-1. Rankit Plot for Original Voice Task Data

Wilk–Shapiro / Rankit Plot of KORG

Approximate Wilk-Shapiro 0.8216   12 cases
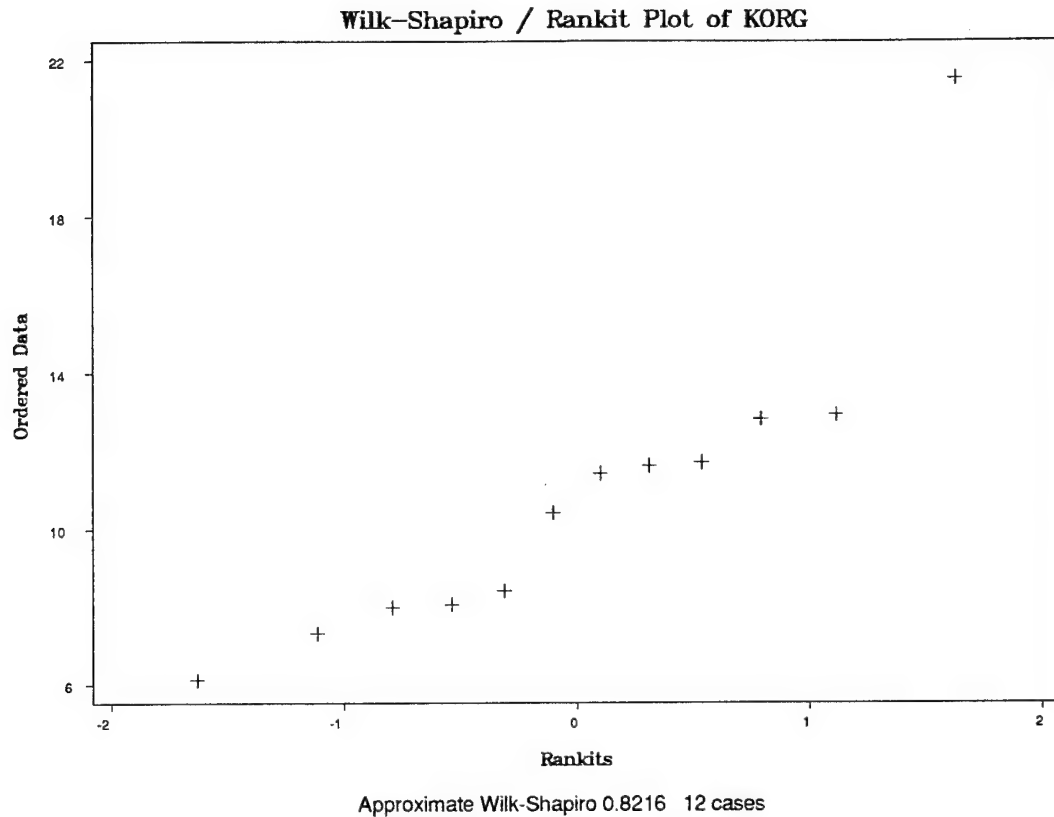
Figure 4-2. Rankit Plot for Original Keypad Data

Discussion of Outlier

The keypad task completion time for subject 1 was significantly greater than the average completion time for all other subjects. It was also significantly greater than the next highest completion time for all other subjects. Further analysis of this subject and task were performed to determine the reason for the extreme difference from the other subjects completion times. During the observation of this subject performing the task, it was noted that the subject used the wrong tool to perform the removal of one of the required components for the task. Rather than using an extended handle hex head screw driver, the subject used a standard allen wrench to remove the item. This sub task was performed in very tight quarters and the use of the improper tool slowed the subject

considerably. This was the first test subject to perform the experiment. All subsequent test subjects were given the extended handle screw driver to remove the item. This is a classic example of finding and using the right tool for the job at hand.

The obvious solution to the problem of statistical outliers is to simply remove the data point from the data set. However, our experimental design precludes us from doing this. Every pair of test subjects perform the experiment according to a Latin square in order to control for learning effect on the hardware. Thus, elimination of a single data point would require the elimination of all data from that pair of test subjects. As we had already exhausted the available pool of personnel to use as test subjects, this was an unacceptable solution. To preclude having to throw out valuable data, an alternative solution was used.

To remove the outlier and not skew the remaining data, the task completion times for subject one were replaced with the task completion times for subject thirteen. Subject thirteen performed the experiment in the same treatment order as subject one. Subject thirteen was able to complete the task, but the data was not originally used in the analysis because of the invalid results from subject fourteen. The adjusted task completion time data are shown in Table 4-2.

A Wilk-Shapiro test for normality was performed on the data shown in Table 4-2. The data for voice task completion times returned a Wilk-Shapiro value of 0.9381. The Rankit Plot is shown in Figure 4-3. Using the minimum value of 0.805, we accept the assumption that this data conforms to a normal distribution. The data for the keypad task completion times returned a Wilk-Shapiro value of .9215. The Rankit Plot is shown in Figure 4-4. This data also conforms to a normal distribution (Conover, 1980).

Table 4-2. Adjusted Task Completion Times

Subject Number	Voice Task Duration	Keypad Task Duration
1	10	13.05
2	12.08	11.62
3	10.02	12.92
4	12.5	8.42
5	7.28	11.42
6	10.85	8
7	8.73	10.42
8	10.83	8.07
9	7.52	12.82
10	12.05	7.33
11	8.63	11.7
12	12.32	6.12
Mean	10.23	10.16
Standard Deviation	1.78	2.34

Wilk–Shapiro / Rankit Plot of VADJ

Approximate Wilk-Shapiro 0.9381   12 cases
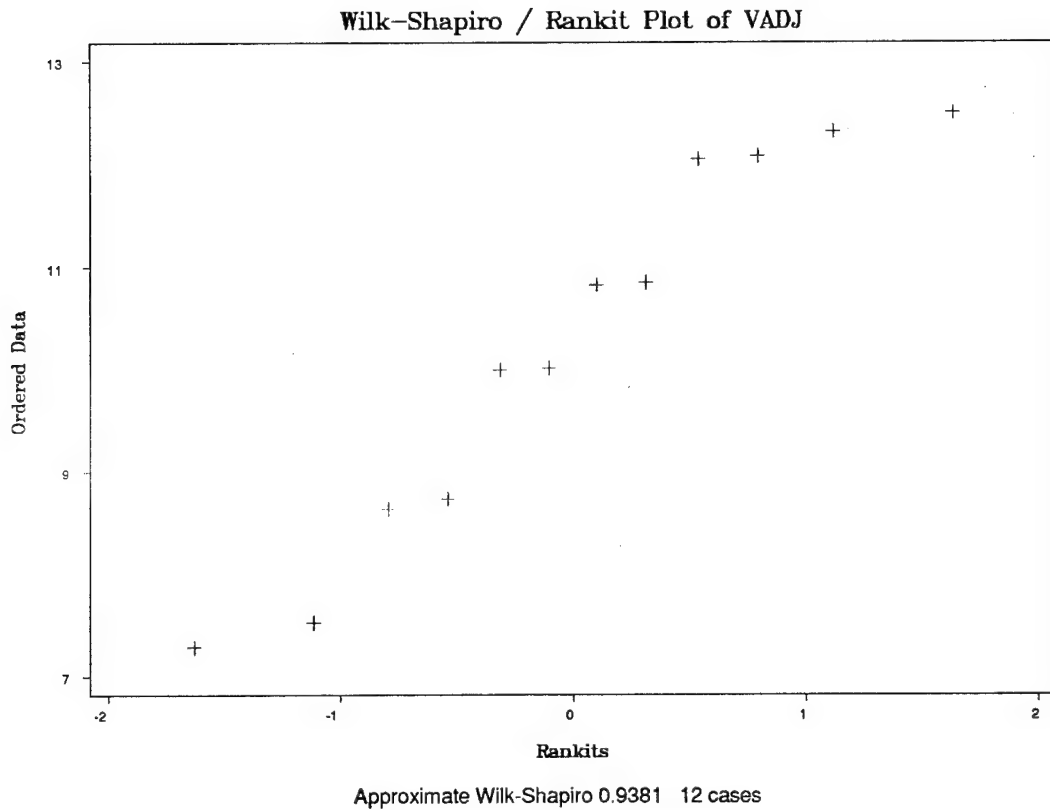
Figure 4-3. Rankit Plot for Adjusted Voice Task Data

Approximate Wilk-Shapiro 0.9215   12 cases
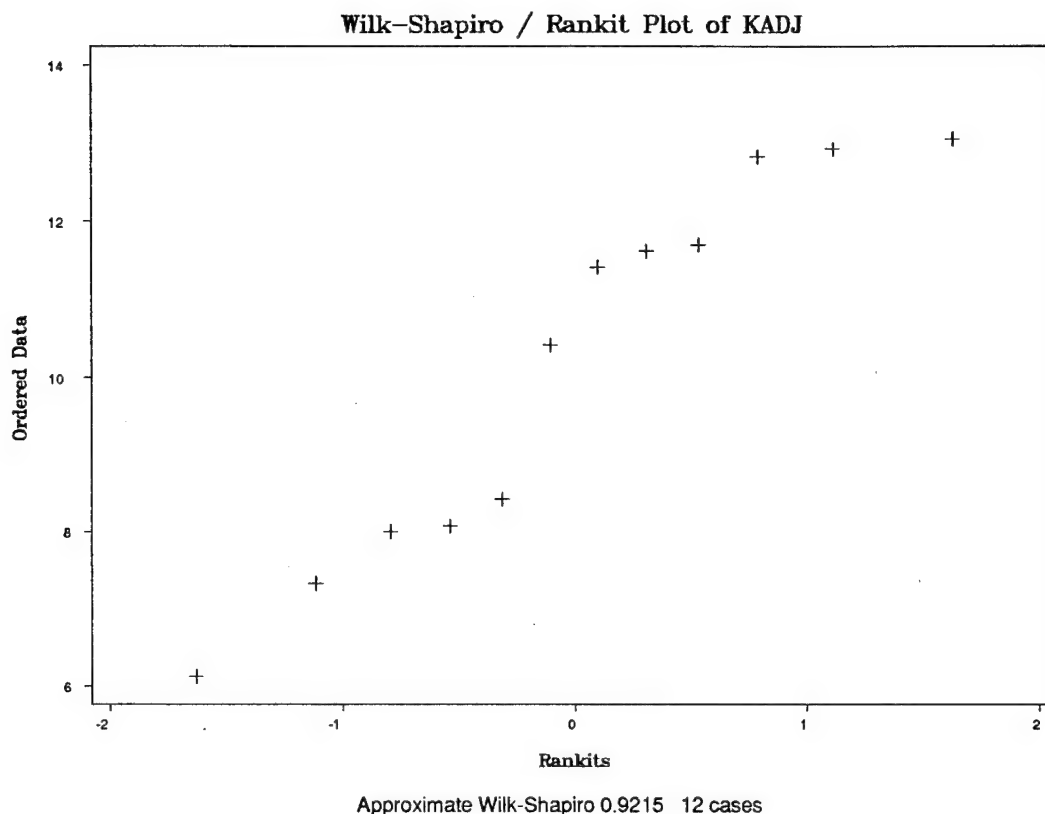
Figure 4-4. Rankit Plot for Adjusted Keypad Data

Paired t-test Results

Having verified the validity and the normality of our data we performed a paired
two sample t-test for means. The calculated t-statistic was 0.06786. The critical t-value
for this data is 1.79588. Comparing these two numbers, we find that the calculated t-
statistic is less than the critical value. Therefore we fail to reject the null hypothesis that
the two means are equal. The observed differences in the means of the two tasks did not
reach conventional levels of statistical significance. The results of this t-test are shown in
Table 4-3.

Table 4-3. Paired t-test Results for Adjusted Data

	Voice Task	Keypad Task
Mean	10.23416667	10.1575
Variance	3.450335606	5.963220455
Observations	12	12
Hypothesized Mean Difference	0	
df	11	
t Stat	0.067863064	
P(T<=t) one-tail	0.473556242	
t Critical one-tail	1.795883691	

Controlling for Computer Processing Time

To control for visible limitations of the voice recognition software, a brief experiment was performed to determine the difference in computer processing times between keypad input and voice input. Keypad input yields a virtually immediate execution of the input command. The voice recognition software package used for this experiment takes a noticeably longer time to execute a spoken command. The computer takes time to record the spoken input and then takes time to compare this recording to the voice templates stored in memory before matching and executing the particular spoken command. To determine what part of the voice task completion times was attributable to computer processing time, we performed a brief experiment. In a laboratory setting, one experimenter executed the commands necessary to simulate the completion of one of the maintenance tasks performed. This simulation followed the exact same path as the test subjects during the field experiment. The simulation was performed ten times using the keypad input and ten times using the voice input. Completion times were recorded using the timing routine on the computer. The mean completion times for both methods were calculated and compared. The difference between the two means was taken as the processing time required to process the voice commands. Results of this are shown in Table 4-4. The mean difference was found to be 48 seconds.

Table 4-4. Table of Times

Run Number	Voice Times	Keypad Times
1	00:01:54	00:01:04
2	00:01:47	00:01:04
3	00:01:50	00:00:58
4	00:01:49	00:01:03
5	00:01:56	00:00:59
6	00:01:45	00:00:59
7	00:01:50	00:01:07
8	00:01:44	00:00:56
9	00:01:45	00:00:56
10	00:01:42	00:00:58
Average	00:01:48	00:01:00

This 48 second compensation factor was subtracted from each individual voice task completion time. This provides a data set of times that compare the input devices on an ideal basis. Resultant data are shown in Table 4-5. Because this compensation was an across the board subtraction for the voice task completion times, the Wilk-Shapiro value did not change. The Rankit plot is identical to the one found in Figure 4-4.

Table 4-5. Compensated Task Times

Subject Number	Voice Task Duration	Keypad Task Duration
1	9.20	13.05
2	11.28	11.62
3	9.22	12.92
4	11.70	8.42
5	6.48	11.42
6	10.05	8
7	7.93	10.42
8	10.03	8.07
9	6.72	12.82
10	11.25	7.33
11	7.83	11.7
12	11.52	6.12
Mean	9.43	10.16
Standard Deviation	1.78	2.34

<u>Paired t-test Results - Compensated Data</u>

Having verified the normality of our data, we again performed a paired two sample t-test for means. The calculated t-statistic was 0.64027. The critical t-value for this data is 1.79588. Comparing these two numbers, we find that the calculated t-statistic is less than the critical value. Therefore we fail to reject the null hypothesis that the two means are equal. The observed differences in the means of the two tasks did not reach conventional levels of statistical significance. The results of the t-test are shown in Table 4-6.

Table 4-6. Paired t-test Results for Compensated Data

	Voice Task	Keypad Task
Mean	9.434166667	10.1575
Variance	3.450335606	5.963220455
Observations	12	12
Hypothesized Mean Difference	0	
df	11	
t Stat	-0.640273253	
P(T<=t) one-tail	0.267552146	
t Critical one-tail	1.795883691	

<u>Summary of Results for Hypothesis I</u>

Results indicate that there is no difference in the mean task completion times between the two input devices. Even when computer processing time for the voice recognition software is controlled for, the mean task completion time difference is not statistically significant. Therefore, Hypothesis I was not supported. Other data collected may indicate that although statistical significance was not achieved, practical significance might be.

Hypothesis II - System Performance

This hypothesis predicted that the system performance with the voice input configuration will meet accepted industry standards. System performance is defined as the

ratio of commands given to commands executed by the computer. For voice recognition software, acceptable levels of system performance are 95 percent recognition rates (Poock and Roland, 1982). System errors are defined as any command that is given and not recognized or improperly executed by the computer. Several system errors were recorded using the voice recognition input device. These errors may be attributable to the influence of unexpected background noise during the experiment. The voice recognition system performed at a 96.81% recognition rate. The individual recognition rates for the voice tasks are shown in Table 4-7.

Table 4-7. Voice Recognition Rates

Subject Number	Voice Recognition Rate
1	100.00%
2	92.50%
3	100.00%
4	94.87%
5	100.00%
6	100.00%
7	97.37%
8	97.37%
9	100.00%
10	97.37%
11	100.00%
12	82.22%
Mean	96.81%

The voice recognition system met the prespecified average performance level, therefore Hypothesis II was supported by the experiment.

Qualitative Results

<u>Hypothesis III - User Satisfaction</u>

This hypothesis predicted that user satisfaction will be greater with the voice input configuration than the keypad input configuration. User evaluation questionnaires were

used to analyze user likes and dislikes of the input devices. Responses to questions about the input devices were measured on a 9-point Likert scale with a middle value of 5. Responses greater than 5 represented a positive response to the question, while responses less than 5 represented a negative response to the question. Open-ended questions were asked about each input device as well as the system in general. A sample of the questionnaires used can be found in Appendices C, D and E.

Keypad Questions

Questions pertaining to the keypad interface were divided into four main subject areas. These areas are overall reactions to the system, learning, system capabilities, and keypad specific questions. Two questions pertained to the overall reaction to the system, two pertained to learning to use the system, three pertained to system capabilities, and three were keypad specific questions. A summary of the responses and means for each question is shown in Table 4-8.

Table 4-8. Summary of Responses for Keypad Questions

Question	Mean
1	7.20
2	7.33
3	8.07
4	7.93
5	7.67
6	7.40
7	6.15
8	8.33
9	8.73
10	8.67

Discussion of Keypad Questions

Question 1. Question 1 evaluated the user's overall reaction to the system with answers ranging from a "terrible" to "wonderful." The mean response was 7.2 which represents a moderate positive reaction to the system.

Question 2. Question 2 evaluated the user's perceived ease of use of the system. Possible answers ranged from "difficult to use" to "easy to use." The mean response was 7.33 which represents a moderately positive perception of the ease of use of the system.

Question 3. Question 3 evaluated the user's perception of how difficult it was to learn how to use the system. Possible answers ranged from "difficult" to "easy." The mean response was 8.07 which represents a positive perception of the difficulty in learning to use the system.

Question 4. Question 4 evaluated the user's perception of difficulty in remembering the location and use of commands. Possible answers ranged from "difficult" to "easy." The mean response was 7.67 which represents a positive perception of the difficulty in remembering the location and use of commands.

Question 5. Question 5 measured the user's perception of the system speed. Possible answers ranged from "too slow" to "fast enough." The mean response was 7.67 which represents a positive perception of the system speed.

Question 6. Question 6 evaluated the user's perception of system reliability. Possible answers ranged from "very unreliable" to "very reliable." The mean response was 7.4 which represents a moderately positive perception of the reliability of the system.

Question 7. Question 7 evaluated the user's perception of the ease of correcting mistakes. Possible answers ranged from "difficult" to "easy." The mean response was 6.15 which represents a barely positive perception of the ease of correcting mistakes.

Question 8. Question 8 evaluated the user's perception of how quickly the system responded to keypad commands. Possible answers ranged from "slow" to "fast." The mean response was 8.33 which represents a positive perception of the speed of system response to keypad commands.

Question 9. Question 9 evaluated the user's perception of how accurately the system responded to keypad commands. Possible answers ranged from "never" to

"always." The mean response was 8.73 which represents a very positive perception of how accurately the system responded to keypad commands.

Question 10. Question 10 evaluated the user's perception of how often the system responded to keypad commands. Possible answers ranged from "never" to "always." The mean response was 8.67 which represents a very positive perception of how often the system responded to keypad commands.

Open-Ended Keypad Questions

There were two open ended questions pertaining specifically to the users likes and dislikes of the keypad. Question one explored what subjects like about the keypad input. Answers focused on the fact that the keypad was quick and easy to use. Several subjects also mentioned the reliability and accuracy of the input. Question two examined what subjects did not like about the keypad input. Answers focused on the fact that the wiring of the keypad to the vest was a limiting factor for the configuration. Several subjects mentioned the difficulty in shifting their focus from the display device and the task to the keypad. Specific answers to these questions can be found in Appendix H.

The qualitative data collected pertaining to the keypad input device reflect a positive acceptance of the keypad as a possible input device for the system as a whole. Subjects appear to be pleased with the system speed, accuracy, and ease of using the keypad input device. Subjects appeared frustrated with the inability to correct mistakes once they were in the middle of a task. The open ended questions verified the acceptance and satisfaction with the speed, accuracy and ease of use of the system. The open ended questions also highlight several limitations of using the current keypad configuration. Although the subjects were pleased with the keypad, several limitations could impair the use of this device. The wiring from the vest is constrictive and the current placement of the keypad may not be optimal.

Voice Questions

Questions pertaining to the voice interface were divided into four main subject areas. These areas are overall reactions to the system, learning, system capabilities, and voice specific questions. Two questions pertained to the overall reaction to the system, two pertained to learning to use the system, three pertained to system capabilities, and three were voice specific questions. A summary of the responses and means for each question is shown in Table 4-9.

Table 4-9. Summary of Responses for Voice Questions

Question	Mean
1	7.00
2	7.13
3	7.60
4	7.33
5	7.33
6	7.33
7	6.08
8	7.27
9	7.73
10	7.80

Discussion of Voice Questions

Question 1. Question 1 evaluated the user's overall reaction to the system with answers ranging from a "terrible" to "wonderful." The mean response was 7.0 which represents a moderate positive reaction to the system.

Question 2. Question 2 evaluated the user's perceived ease of use of the system. Possible answers ranged from "difficult to use" to "easy to use." The mean response was 7.13 which represents a moderately positive perception of the ease of use of the system.

Question 3. Question 3 evaluated the user's perception of how difficult it was to learn how to use the system. Possible answers ranged from "difficult" to "easy." The

mean response was 7.6 which represents a positive perception of the difficulty in learning to use the system.

Question 4. Question 4 evaluated the user's perception of difficulty in remembering the names and use of commands. Possible answers ranged from "difficult" to "easy." The mean response was 7.33 which represents a moderately positive perception of the difficulty in remembering the names and use of commands.

Question 5. Question 5 measured the user's perception of the system speed. Possible answers ranged from "too slow" to "fast enough." The mean response was 7.33 which represents a moderately positive perception of the system speed.

Question 6. Question 6 evaluated the user's perception of system reliability. Possible answers ranged from "very unreliable" to "very reliable." The mean response was 7.33 which represents a moderately positive perception of the reliability of the system.

Question 7. Question 7 evaluated the user's perception of the ease of correcting mistakes. Possible answers ranged from "difficult" to "easy." The mean response was 6.08 which represents a barely positive perception of the ease of correcting mistakes.

Question 8. Question 8 evaluated the user's perception of how quickly the system responded to voice commands. Possible answers ranged from "slow" to "fast." The mean response was 7.27 which represents a moderately positive perception of the speed of system response to keypad commands.

Question 9. Question 9 evaluated the user's perception of how accurately the system responded to voice commands. Possible answers ranged from "never" to "always." The mean response was 7.73 which represents a positive perception of how accurately the system responded to keypad commands.

Question 10. Question 10 evaluated the user's perception of how often the system responded to voice commands. Possible answers ranged from "never" to "always." The

mean response was 7.8 which represents a positive perception of how often the system responded to keypad commands.

Open-Ended Voice Questions

There were two open ended questions pertaining specifically to the users likes and dislikes of the voice recognition system. Question one explored what subjects like about the voice input. The overwhelming response to this question focused on the fact that the voice system provided a hands free environment for the technician. Several subjects also mentioned the positive aspects of not having to shift their focus away from the tech data and the task at hand. Question two examined what the subjects did not like about the voice input. Answers focused on the limitations of the voice recognition software. Specifically, subjects mentioned problems encountered dealing with background noise and the inability of the software to distinguish between normal speech and command inputs. Specific answers to these questions can be found in Appendix I.

The qualitative data collected pertaining to the voice input device reflect a positive acceptance of the voice recognition system as a possible input device for the system as a whole. Subjects appear to be pleased with the system accuracy, and ease of using the voice input device. Again subjects appeared frustrated with the inability to correct mistakes once they were in the middle of a task. The open ended questions reveal a perceived advantage using voice as opposed to the keypad because of the hands free capability provided by the voice system. The open ended questions also highlight several major limitations of the software package.

Overall Questions

There were seven open ended questions pertaining to the overall evaluation of the system. These data were collected to aid in determining user preference and to provide positive feedback for potential system improvements.

Question 1. *What did you like about the image in the eye-piece?* Subject responses to this question were generally positive. Subjects commented on the positive aspects of having tech data constantly available and not having to carry, hold or refer to paper tech manuals while performing a task.

Question 2. *What did you not like about the image in the eye-piece?* Subject responses to this question centered on problems encountered focusing properly on the eye-piece. Specifically, subjects mentioned the size of the image and the lack of color in the display. Subjects also commented on the fact that the eye piece moved occasionally during the task.

Question 3. *What did you like about the "suite" of eye-piece, vest, input devices, etc..?* Responses to this question focused on the positive aspect of the mobility provided by the vest configuration. Subjects spoke highly of the ability to move around the airplane and still have access to tech data. One subject had nothing positive to say about the suite and one subject said the configuration should be trimmed down.

Question 4. *What did you not like about the "suite" of eye-piece, vest, input devices, etc..?* The overwhelming majority of subjects commented that the vest was restrictive and bulky. Several subjects noted that access to several confined areas of the aircraft would be restricted.

Question 5. *Which input device would you prefer to use on a daily basis?* Eight subjects commented that they would prefer to use the voice input. Four subjects said they would prefer to use the keypad input. Two subjects mentioned no preference. Two subjects did not have the opportunity to sufficiently evaluate both input devices.

Question 6. *Do you have any suggestions for future study?* Subjects suggested testing a smaller version of the vest configuration. Several subjects also mentioned performing a test to evaluate how well schematics and diagrams can be viewed while using the system.

Question 7. *Do you have any suggestions for future hardware improvements?* Responses focus on the miniaturization of the hardware, weight reductions to the hardware, and improvements to the eye-piece.

The qualitative data collected in the open ended questions about the system in general basically say that the users do not reject the system for use in performing maintenance tasks. However, they do point out several things to improve both the performance of the system and the acceptance by the users. This data also reflects a preference for the voice input device over the keypad input device. This is primarily due to the hands free environment provided by the voice system. Specific answers to these questions can be found in Appendix J.

<u>Summary of Results for Hypothesis III</u>

Results indicate that subjects prefer to use the voice input device over the keypad input device. Even though the qualitative data show the two devices to be nearly equal, and the voice recognition software has some limitations, subjects still preferred the voice input over the keypad. The determining factor appears to be the hands free capability provided by the voice input device.

Summary

Three types of data were collected for this experiment: task completion times, system performance results, and user responses to input devices and overall system. Task times were recorded for each voice task and each keypad task. Times were recorded using a timing routine installed on the computer system and then verified using times recorded by the observer using a stopwatch. In addition to task completion times, the number of commands not recognized by the computer were recorded for each voice and keypad task. Qualitative evaluations were performed by each test subject after each task performed. Each subject answered 10 questions evaluating each input device. Questions

were developed using a nine point Likert scale. Open-ended questions were also answered on each input device and the overall system. Experimental Hypothesis I was not supported by the data collected. Experimental Hypotheses II and III were supported by the data collected. Valuable subjective information was gathered during this experiment.

V. Discussion, Recommendations, and Conclusions

Chapter Overview

This chapter contains a discussion of the results found during this experiment. A discussion of quantitative and qualitative data addresses the support found for our experimental hypotheses. Recommendations for system improvements noted by experimental subjects as well as the experimenters are discussed. Recommendations for further research are presented. Conclusions drawn from this experiment are presented.

Discussion of Quantitative Results

Quantitative data were collected for evaluation of the first two experimental hypotheses. Task completion time data was collected to evaluate Hypothesis I and system performance data was collected to evaluate Hypothesis II.

Hypothesis I

The data collected during this experiment does not support experimental Hypothesis I. This hypothesis states that task completion times using voice recognition will be faster than task completion times using the keypad. Statistical significance was not reached in the comparison of the mean task completion times for each input device. We believe there are several factors contributing to the lack of statistical significance of the results. These factors are the type of task evaluated, the type of user evaluated, and the learning effect associated with using the system.

Type of Task

The tasks that were used in this experiment required the technicians to take a maximum of six steps to evaluate the fault and identify the corrective action. These steps were relatively easy to perform and therefore did not take a very large amount of time. There was also very little movement required by the technician which also would have

added time to complete the required task. While this task is representative of the routine maintenance performed during avionics troubleshooting, it may not be indicative of the number of steps required in many other troubleshooting tasks.

Another point to consider is the fact that there were additional constraints placed on the experiment by the owning organization of the test-bed aircraft. These constraints include availability of the aircraft, availability of personnel, and the amount of repeated maintenance allowed by the owning organization. The Springfield ANG unit is an operational unit that has a daily flying commitment. We were limited to using aircraft that were not required to meet the flying schedule. Additionally, personnel support was limited by the number of technicians in the organization who were not working priority maintenance issues.

Whenever maintenance is performed on an aircraft, there is always the danger of damaging a serviceable asset. In a repeated maintenance task this danger is greatly increased. In an effort to minimize this risk, several precautions were taken. The aircraft was made safe for maintenance to preclude having each technician climb into the cockpit before each task. All panels required to be opened for the performance of the tasks were opened ahead of time, removing the possibility of the technicians stripping the screws on the panels. The line replaceable units required to be removed during the task were removed ahead of time, which helped avoid any damage to the aircraft or unit during repeated removal and installation. These factors produced an over simplification problem by reducing the total number of steps required to complete the tasks and therefore reducing the amount of time required to complete the tasks.

Type of Users Evaluated

Avionics maintenance technicians were used as subjects in this experiment. These technicians typically do not perform tasks that require great amounts of movement around the aircraft or frequent use of both hands. Perhaps more definitive results would be found

using technicians and tasks from a different maintenance specialty to perform the experiment. For instance, a weapons load task would require constant movement around the aircraft as well as extensive use of both hands. A load task would also require a constant referral to a checklist. Another possibility is the use of crew chiefs to perform a preflight inspection or a tire change using the experimental equipment. Making the task more stringent will better highlight the differences between the voice and keypad input devices. A longer and more complex task may yield statistically significant results.

Learning Effect

One important point to be noted from these experimental results is the fact that there was a significant learning effect shown by each technician. No matter which input device they used first, the device used second produced much faster results. While our experimental design was built to control for this learning effect, this control produced variability in the results. The inclusion of more training prior to starting the experimental tasks may reduce the learning effect and therefore reduce the variability found in the results.

Hypothesis II

The data collected during this experiment supports experimental Hypothesis II. This hypothesis states that the voice recognition software package will perform within acceptable industry standards. Performance data viewed from a macro perspective shows that the system operated with an average 96.8 percent recognition rate. This fact supports the experimental hypothesis. There are several factors contributing to the support of this hypothesis. These factors include that the experiment was performed in a relatively noise free environment, the conditions under which the voice package was trained, and the fact that the software package is a speaker dependent program. Closer examination of the experimental data reveals that several test subjects experienced below standard performance of the software package. The reasons for the substandard performance can

be traced to periods of uncontrolled high noise levels for which the voice package was not trained.

Experimental Environment

The experiment was performed inside an aircraft maintenance hangar. This helped to limit several noise factors. The technicians were isolated from the noise of an active aircraft flight line. This includes noises associated with aircraft taxiing, ground equipment operating, vehicle traffic, and yelling production supervisors and squadron maintenance officers. The technicians were also not required to have any ground power equipment operating as a part of the task. There were several instances where noise inside the hangar could not be controlled. These instances account for the substandard system performance rates. For example, during two voice tasks, a hydraulic test stand was started and operated in close proximity to the test bed aircraft. The voice package did not perform well during these two tasks. As long as we were able to control for all of the noise in the experimental environment, the voice recognition package responded well.

Training

Another factor that had a positive influence on the performance of the voice package was the training of the voice recognition software. There are two considerations involved in training, the environment in which the package is trained and the number of times each command is repeated during training. Through trial and error we found that the environment in which the software is trained had a significant effect on the accuracy of the recognition. Higher recognition rates were achieved when the software was trained in noise levels close to those in which the task was to be performed. For example, one voice task was completed with a hydraulic test stand running close to the technician. The voice package was trained in that noise filled environment and then used in the same environment. The recognition rate was 100%. Conversely, another voice task was completed where the voice package was trained in a quiet environment. Noise levels

increased part way through the task. In this scenario, a recognition rate of only 82% was achieved.

The number of times a command was repeated during training also had an impact on the accuracy of the system. Once again through trial and error, before the experiment was begun, it was determined that training each command three times led to recognition rates between 90 and 95 percent. Increasing the number of repetitions to five, improved recognition rates to greater than 95%. Increasing the number of repetitions above five resulted in no significant improvement in recognition rates. Therefore, by using five repetitions we were able to achieve satisfactory performance while adding minimal time to the voice training requirements.

Speaker Dependent Software Package

Another factor influencing the positive system performance is the fact that the software package used for this experiment was a speaker dependent system. In a speaker dependent system, each user trains the computer to recognize the patterns of his voice. Speaker dependent systems are generally recognized as more accurate than speaker independent systems. Speaker independent systems have a broad voice template already established. The user's voice is compared to this broad template until a match is found. The variability of user's voices may have a negative impact on the reliability of the recognition package. Because the template must be so broad, this leaves more room for error. While the speaker dependent systems require additional time for training, the improved accuracy provided more than makes up for the extra time investment.

Discussion of Qualitative Results

Qualitative data were collected for evaluation of the third experimental hypothesis. User preference information was collected to evaluate Hypothesis III.

Hypothesis III

Hypothesis III states that user satisfaction will be greater with the voice input configuration than with the keypad configuration. Results of the qualitative data collected show that test subjects preferred the voice input configuration over the keypad. There were two reasons for this preference. Subjects liked the hands free capability provided by the voice input and they liked the fact that they did not have to shift their focus away from the task to input commands to the computer.

Hands Free Capability

Normal aircraft maintenance requires technicians to perform tasks while constantly referring to possibly several volumes of tech data at one time. This constant referral requires the technician to remove his hands from the task in order to find and follow the necessary steps required to complete the task. The use of voice recognition eliminates the need for the technician to remove his hands from the task to access the information necessary to complete the job. This feature was found to be overwhelmingly accepted and appreciated by the technicians performing this experiment.

Shift of Focus

Another distinguishing factor noted by our test subjects was the ability to remain focused on the tech data or task when inputting commands to the computer. Even though the data were constantly available with the keypad, the technician was forced to shift his focus to the keypad to input commands to the computer. This was highlighted by the test subjects as a distinguishing feature that made the voice input device more favorable.

Improvements

Recommendations for system improvements were noted by both experimental subjects and the experimenters. These suggestions for improvement cover four main

areas, the eyepiece, the voice package, the keypad, and the suite in general. Discussion of the suggestions follows.

Eyepiece

The monocular head mounted display device used during this experiment was generally accepted by the test subjects. However, numerous suggestions were made that would improve the user satisfaction with the equipment. The ideal display device envisioned by the test subjects was a heads-up type device that you could look through rather than the current picture tube type that you cannot see through. The ideal device would be a binocular or visor type device that presents an image that can be viewed both eyes. The device would be small and light weight similar to a pair of glasses. These glasses would be comfortable and small enough to allow access to confined areas. Ideally these glasses could be worn in a chemical environment underneath a gas mask. We realize these are stringent requests, but similar devices are being manufactured in industry today.

Voice

Several improvements to the voice recognition package are necessary before voice becomes a truly viable option as a user interface for the flight line maintenance technician. These improvements focus on speed and noise filtering. There are commercially available voice recognition packages that will more quickly process and execute a spoken command. Noise filtering can be accomplished through two methods. Enclosing the microphone in a manner similar to the communications headsets currently used by maintenance technicians would be a possible solution to the problem of noise elimination. The development of high frequency, high decibel noise filters is a necessity for use around turbine engines found in airplanes and in ground power equipment.

Keypad

If the keypad is to be pursued as a potential user interface, several modifications should be considered. The use of the keypad could be improved by eliminating or re-

5-7

routing the wiring from the keypad itself to the computer. Ideally, a wireless keypad may yield the greatest benefit. Additional consideration needs to be given to the optimal placement of the keypad. Wearing the keypad on the wrist limits the ability of the user to reach into small areas. This also detracts from the hands free environment that the equipment can provide. Another improvement to the keypad would be the use of recessed keys. The buttons on the current keypad can be inadvertently pushed giving incorrect and erroneous command input. Closely related to the redesign of the keys is the examination of the number of commands required to fully manipulate the system. Use of only the essential commands may allow for the elimination of several keys which could reduce the size of the required keypad.

Suite

There are several improvements that should be made to the suite in general. These improvements are not aimed specifically at either input device. The size and weight of the components should be targeted for reduction. Ideally, all components should be small enough to be placed together and worn comfortably on a belt. This would allow for easier movement and greater mobility in and around confined areas. Currently the system is designed to be powered by two or three camcorder batteries with a three hour battery life limitation. Considering that this system will be used during an extended working day, the current power supply is insufficient. There are several alternatives to counter this problem. Batteries that last longer and weigh less would improve the practicality of the system. Other solutions include adapting the system to be powered by aircraft power and/or ground power units. This would help to overcome the battery life limitation but would detract from the mobility of the system. Another improvement that must be considered is the integration of the system into the aircraft communications system. Many flight line tasks that must be performed require the use of the aircraft intercom. To

effectively use the capability of this system, it must work in coordination with the aircraft communication system.

Recommendations for Further Research

In the performance of this experiment it became obvious that more research would be beneficial before a final system configuration is agreed upon for production. Further research may identify other avenues to further improve and refine the system. We believe that voice recognition should be pursued as the input device of choice. Further study should evaluate the voice recognition capability, as well as the suite in general, in a more stringent environment. Additional study may highlight more significant advantages gained by using voice over any other type of input device. Task selection for any future experiment should consider both the type of task and the noise environment in which the task will be performed.

When choosing the type of task to be used in the experiment, several factors should be considered. The complexity of the task, the degree to which tech data must be used, and the degree to which both hands are used to complete the task should drive the selection. A more complex task that includes more steps, requires more movement around the aircraft, and requires the use of more equipment may better evaluate the practical advantages provided by the system. A more complex task may require technicians to rely more heavily on the tech data. A more complex task may also require a greater use of both hands to complete the task. For example, a weapons load task, a hydraulic pump change, a flight control troubleshooting task, or a tire change would be excellent tasks to evaluate the effectiveness of the system.

When choosing the environment in which to perform the task, consideration should be given to the normal environment in which the task will be performed. If the task chosen is a flight line task, then normally experienced flight line noise should be included.

In order to more effectively evaluate the voice recognition capability, experimentation needs to be performed in a more realistic or noisier environment. In our experiment, because of the constraints placed on us by the supporting organization, we did not test the voice recognition package in a worst case environment. Further study should be accomplished to ensure that the software package can fully function in the environment in which it will be used.

Conclusion

The voice recognition input device provided features that the test subjects found most desirable in the aircraft maintenance environment. The hands free capability and the ability to remain focused provided by the voice recognition system well outweighed the slower computer speed and the slightly lower accuracy of command input. The fact that technicians preferred the voice system over the keypad system even though there was no statistical difference in performance is very important. In a metaanalysis of usability studies, Nielson and Levy concluded that users opinions and preferences are valuable data and should be taken into account when choosing between user interface designs. Their research indicates that there is a strong positive association between users average task performance and their average subjective satisfaction. There is a reasonably large chance for success if the selection of user interface is based solely on users opinions (Nielson and Levy, 1994). Based on the results of this experiment it would appear that the voice interface should be pursued as the interface of choice for the integrated maintenance information system.

Appendix A. Experimental Plan

I. Description of Evaluation

Purpose

The purpose of this study is to evaluate performance differences between technicians performing tasks using voice input versus keypad input to manipulate technical information presented on a HMDD.

Hardware

A single piece of equipment will be used in this experiment. This will consist of a vest-mounted computer used to display digital technical data on a HMDD. The vest will be equipped with a voice recognition input device and a keypad input device.

Software

Software for this experiment includes PCIMIS and VoiceAssist. PCIMIS is a Windows-driven presentation system used for presenting digital technical data. The voice recognition software is VoiceAssist, developed by Creative Labs, Inc.

Subjects

There will be a total of 16 maintenance technicians from the 178 TFG participating in this experiment. Technicians will be divided into two groups. One group will accomplish the voice task first and the other group will accomplish the keypad task first.

Tasks

Each maintenance technician will complete two different troubleshooting tasks. One task will be performed using the voice input device and the other task will be completed using the keypad input device. Table A-1 shows the experimental conditions.

Table A-1. Experimental Conditions

	Voice Input	Keypad Input
Group 1	Perform task 1 first	Perform task 2 second
Group 2	Perform task 2 first	Perform task 1 second

The troubleshooting tasks have been determined to be typical maintenance tasks and parallel based on the time required for completion and the degree of difficulty. The tasks will require the technician to isolate a faulty component in the F-16C/D Heads-Up Display system.

Conditions

The task and input device will be counterbalanced to control for order and learning effects. All subjects will be randomly assigned to their experimental conditions.

Hypotheses

Hypothesis I predicts that the task completion times using the voice input device will be faster than the task completion times using the keypad input device.

Hypothesis II predicts that the system performance with the voice input device will meet accepted industry standards.

Hypothesis III predicts that user satisfaction will be greater with the voice input device than with the keypad input device.

Data Collection

Maintenance experience data will be collected using the Personal Background Form (Appendix B). The information from this form will be used to draw similarities between Guard and Active Duty training requirements. During the experiment, notes, observations, and task start and stop times will be documented by the experimenter on note paper. A questionnaire will be administered after the use of each input device and after each subject has completed both tasks (Appendices C, D, E). The following information will be collected during the experiment:

- Task completion time
- Command input failures
- User preference data

Controls

The following actions will be performed to control for experimental variation.

1. The same experimenters will conduct all of the data collection activities.

2. All data collection runs will be performed at the same location.

3. All subjects will be randomly assigned to one of two experimental groups.

4. All subjects will be tested for 20/20 corrected or uncorrected vision.

5. Subjects will be asked to perform a test to determine their dominant eye.

6. Each test subject will receive identical training on how to use the system, as well as how to use each input device.

7. Computer response times for the voice recognition software will be determined and subtracted from the voice task completion times.

8. System errors, including both maintenance and computer errors, will be handled identically. If an error is made, the task will be halted and re-started at the same place the error occurred.

9. Learning effect will be controlled for by alternating which input device is used first by each test subject. Half of the test subjects will use the voice input first and the other half will use the keypad input first.

10. Order effect will be controlled for by alternating which maintenance task is performed first by each test subject. Half of the subjects will complete Task One first and the other half will complete Task Two first.

III. Conducting the Experiment

Sequence of Events

- Eye examinations and eye dominance tests will be performed (minimum is 20/20 with corrected vision)
- Demonstration of the IMIS system (from the laptop) to the subjects hands-on computer training (with eye piece) technician 1
- Technician 1 to perform experimental condition 1, Technician 2 perform hands on computer training (with eye piece)
- Technician 2 to perform experimental condition 2, Technician 1 completes questionnaire and has familiarization training with eyepiece
- Technician 1 to perform experimental condition 3, Technician 2 completes questionnaire and has familiarization training with eyepiece
- Technician 2 to perform experimental condition 4, Technician 1 is debriefed (to include the final questionnaire)
- Technician 2 is debriefed (to include the final questionnaire)

Introduction

Before the experiment begins, the technicians will be required to read and sign the Human Use Release Form shown in Appendix K. The technicians will then receive a briefing of the purpose of the experiment and the instructions required for completion of the experiment. The briefing will cover voluntary participation, background of the research, what the technicians will be required to do, the data that is to be collected, and

the privacy of their performance and responses. Technicians will also be given the information necessary to begin and complete each task. The briefing instructions are provided in Appendix F.

Training

Technicians will receive training in two areas. First, training will be provided on the operation of PCIMIS and the commands required to perform required actions. Second, each technician will receive training on each input device immediately before it is used to perform the task. The experimenters will be available at all times during the training to answer any questions the technicians may have.

Debriefing

Questionnaires will be administered to each test subject after completion of each task. Additionally, a questionnaire will be administered after both tasks have been completed. These questionnaires can be found in Appendices C, D, and E. After the test subject completes the final questionnaire, he/she will be debriefed. Debriefing will include asking the participant for any other feedback, thanking him/her for participating, and requesting that the test subjects not discuss the experiment with anyone until the experiment has been completed. The debriefing instructions are provided in Appendix G.

Appendix B: Personal Background Form

Personal Background

Name: _____

1. Check one: Full - Time   Traditional
                  Guardsman _____   Guardsman _____

2. Time in Service: _____ (yrs / mos)

3. Paygrade / Rank: _____

4. Current Specialty: _____ (Job Title / AFSC)

    Any others held: _____

5. Prior work experience:
                          (1)   (2)   (3)   (4)
    Weapons System _____   _____   _____   _____
    Number of Years _____   _____   _____   _____

6. Education / Training:
    How many Career Development Course volumes have you completed?
    5 level: 2A532   2 volumes   _____
                  A shreddout 4 volumes   _____
    7 level: A shreddout 4 volumes   _____
                  B shreddout 2 volumes   _____
                  C shreddout 3 volumes   _____
    What Avionics Maintenance FTD courses have you attended?
            Communication / Navigation Course   _____
            Flight Controls Course   _____
            Comm / Nav and Penetration Aids Course _____
            Attack Control Course   _____
    Rate your experience level with the HUD system

Inexperienced	Somewhat Inexperienced	Average Experience	Somewhat Experienced	Experienced
+--------------------	+--------------------	+--------------------	+--------------------	+

7. Computer experience:
   Rate your computer experience level

Inexperienced	Somewhat Inexperienced	Average Experience	Somewhat Experienced	Experienced

+----------------------+----------------------+----------------------+----------------------+
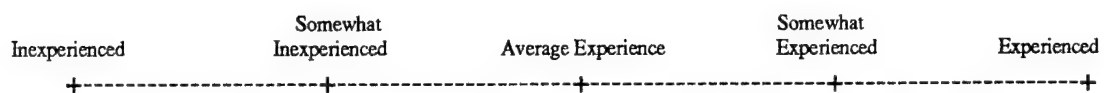
   Are you familiar with Microsoft Windows? _____

Appendix C. Keypad Questionnaire

Qualitative Data Collection - Keypad Interface

Subj No._____

Overall reactions to the system

Terrible							Wonderful	
1	2	3	4	5	6	7	8	9

Difficult to use							Easy to use	
1	2	3	4	5	6	7	8	9

Learning

Learning to operate the system

Difficult								Easy
1	2	3	4	5	6	7	8	9

Remembering location and use of commands

Difficult								Easy
1	2	3	4	5	6	7	8	9

System Capabilities

System speed

Too slow							Fast Enough	
1	2	3	4	5	6	7	8	9

How reliable is the system?

Very Unreliable							Very Reliable	
1	2	3	4	5	6	7	8	9

Correcting your mistakes

Difficult								Easy
1	2	3	4	5	6	7	8	9

Keypad

How quick did the system respond to keypad commands?

Slow								Fast
1	2	3	4	5	6	7	8	9

How accurately did the system respond to the keypad commands?

Never   Always

1   2   3   4   5   6   7   8   9

How often did the system respond to keypad commands?

Never   Always

1   2   3   4   5   6   7   8   9

Open Ended

What did you like about keypad input?

What did you not like about keypad input?

Appendix D. Voice Questionnaire

Qualitative Data Collection - Voice Interface

Subj No._____

Overall reactions to the system

Terrible   Wonderful
1   2   3   4   5   6   7   8   9

Difficult to use   Easy to use
1   2   3   4   5   6   7   8   9

Learning

Learning to operate the system
Difficult   Easy
1   2   3   4   5   6   7   8   9

Remembering location and use of commands
Difficult   Easy
1   2   3   4   5   6   7   8   9

System Capabilities

System speed
Too slow   Fast Enough
1   2   3   4   5   6   7   8   9

How reliable is the system?
Very Unreliable   Very Reliable
1   2   3   4   5   6   7   8   9

Correcting your mistakes
Difficult   Easy
1   2   3   4   5   6   7   8   9

Voice

How quick did the system respond to voice commands?
Slow   Fast
1   2   3   4   5   6   7   8   9

How accurately did the system respond to the voice commands?
Never   Always
1   2   3   4   5   6   7   8   9

How often did the system respond to voice commands?
Never   Always
1   2   3   4   5   6   7   8   9

Open Ended

What did you like about voice input?

What did you not like about voice input?

Appendix E. Overall System Questionnaire

Qualitative Data Collection - Final

Subj No._____

Open Ended

What did you like about the image in the eye-piece?

What did you not like about the image on the eye-piece?

What did you like about the "suite" of eye-piece, vest, input devices, etc..?

What did you not like about the "suite" of eye-piece, vest, input devices, etc..?

Which input device would you prefer to use on a daily basis?

Do you have any suggestions for future study?

Do you have any suggestions for future hardware improvements?

Appendix F. Briefing Instructions

Introduction

Thank you for volunteering to be a test subject in the evaluation of input devices for the manipulation of digital maintenance technical data.

I am Captain Dave Chapman, and this is Captain Jim Simmons. We are graduate students at the Air Force Institute of Technology at Wright-Patterson AFB and we are performing this study as part of our master's thesis requirement.

Purpose

The objective of the evaluation is to compare voice input versus keypad input for manipulation of digital maintenance technical data. (Point to HMDD)

This study is sponsored by the Armstrong Laboratory, located at Wright-Patterson Air Force Base. The study is part of a program at the Laboratory, titled the Integrated Maintenance Information System (IMIS). IMIS is a development project aimed at providing technicians with a flight line computer to support maintenance activities. IMIS will contain technical data for the aircraft, tie into the supply computer for spare parts information, provide automated diagnostic routines, display historical information either from CAMS or the wing, and will tie into unit training requirements. The information obtained from this experiment will support the development of the IMIS user interface technologies.

Experimental Description

There are a total of 16 maintenance technicians participating in this experiment. You will receive training on how to use each of the two input devices (voice and keypad). Then you will perform two maintenance tasks, one using one device, and the other using the second device. First, you will be asked to perform a task using either the voice input or the keypad input. Your job will be to perform the steps displayed on the eye-piece and

isolate any faults or discrepancies you encounter. We will be recording the total task completion time and errors made using the input device. Therefore, you are encouraged to work as quickly and accurately as possible. After completing the first task, you will be given a break, then you will perform the second task using the other input device. After you complete each phase of the experiment, you will be asked to complete a questionnaire which addresses certain aspects of the user interface and the information displayed. Remember the input devices are being compared in this study, you are not being studied. The information collected in this experiment will not be associated with your name. Participants will only be identified by subject number. The data collected will not be related to your job performance. Your supervisor will not know how you did on the experiment nor will he hear any of the comments you provide during the debriefing. The sequence of events will be as follows:

- eye examinations and eye dominance tests will be performed (minimum is 20/20 with corrected vision)
- demonstration of the IMIS system (from the laptop) to the subjects
- hands-on computer training (with eye piece) Technician 1
- Technician 1 to perform experimental condition 1, Technician 2 perform hands on computer training (with eye piece)
- Technician 2 to perform experimental condition 2, Technician 1 completes questionnaire and has familiarization training with eyepiece
- Technician 1 to perform experimental condition 3, Technician 2 completes questionnaire and has familiarization training with eyepiece
- Technician 2 to perform experimental condition 4, Technician 1 is debriefed (to include the final questionnaire)
- Technician 2 is debriefed (to include the final questionnaire)

Instructions

You will be provided with the equipment necessary to complete the task. A toolbox, multimeter, jumper wires, and magnifying glass will be available. We have inserted a fault into the system for each task using a breakout box connected to one of the LRUs. You can assume that the plug connected to the LRU is the aircraft wiring. In the tech data, remember that J indicates a jack and P indicates a plug. You will be required to perform an ohms check on a cannon plug and a wafer plug. Pin HH is at the center of the cannon plug. The wafer jack is numbered from left to right. The top row is row A and the bottom row is row B. If you need to remove/replace parts during the task, simulate actual removal/replacement and proceed with the task. Some parts have already been removed from the task. If the tech data asks you to remove a unit, just proceed with the task. You can assume that the aircraft has been made safe for maintenance.

Appendix G. Debriefing Instructions

Thank you for participating in this evaluation. The purpose of the experiment was to compare troubleshooting performance with digital technical data being manipulated with a keyboard interface to performance with data being manipulated with voice input. The information from this evaluation will support the selection of the user interface with this digital information.

None of the information received or data collected will be associated with your name. Experimental write-ups will describe the data only by subject number. Do you have any other comments about the evaluation?

Thanks again for your participation. We would appreciate your not discussing any aspect of this evaluation with your co-workers until all of the data has been collected.

Appendix H. Responses to Keypad Questions

Responses to Keypad Questions

Subject Number	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	6	3	8	7	9	8	8	9	9	8
2	7	7	7	7	7	7	7	7	7	9
3	7	7	9	9	7	7	1	7	9	9
4	7	8	8	8	8	7		9	9	9
5	9	8	9	9	8	8		9	9	9
6	8	8	8	8	8	8	7	8	8	8
7	7	7	8	8	6	7	7	9	9	9
8	4	6	7	8	7	7	8	8	9	9
9	8	8	9	9	8	9	8	8	9	8
10	8	7	7	7	8	8	5	9	9	8
11	7	8	8	7	6	6	7	7	9	9
12	8	8	8	8	8	8	8	8	8	8
13	6	8	7	7	7	5	7	9	9	9
14	7	8	9	9	9	7	6	9	9	9
15	9	9	9	8	9	9	1	9	9	9
16										
Mean	7.20	7.33	8.07	7.93	7.67	7.40	6.15	8.33	8.73	8.67
Standard Deviation	1.22	1.35	0.77	0.77	0.94	1.02	2.35	0.79	0.57	0.47

Open-ended Questions

Question 1 - What did you like about keypad input?

Subject 1
        quick in responding to inputs
Subject 2
        no noise interference from outside sources
Subject 3
        keypad easy to use
Subject 4
        it was easy to use
Subject 5
        Simple to use. Reliable input.
Subject 6
        I felt like I could manipulate the system faster and the system always
responded quickly and correctly as long as I pressed the right buttons.

Subject 7
    The keypad was small and easy to use. It did not get in the way.
Subject 8
    Better than the voice command because of the fact that you can work and talk with another person
Subject 9
    Simple, which is imperative with maintenance personnel
Subject 10
    The keypad was faster. I felt more comfortable with the keypad. I didn't need to "think" of what commands to say. I knew what the function of the keys were.
Subject 11
    Very few buttons to deal with.
Subject 12
    It was very quick and accurate.
Subject 13
    It worked good. I think it is a great idea.
Subject 14
    not dependent to voice recognition. Works in a noisy environment.
Subject 15
    Easy access to tech data not currently being used. Can re-read instructions while in the middle of task with out leaving immediate job site
Subject 16
    no response

Question 2 - What did you not like about keypad input?

Subject 1
    the wiring going to the keypad makes performing a task more difficult
Subject 2
    did not have total freedom on hands - but wasn't bad
Subject 3
    not enough functions, wired to computer, need to change pointing device
Subject 4
    I liked it
Subject 5
    Wire harness (for keypad) slightly interfered with arm actions.
Subject 6
    I think it functioned very well but there was even less mobility because as I was reaching to put a tool down with the hand that was connected to the keypad I didn't have the full range or extension ability of my arm.
Subject 7
    The wires attached might get caught up or get ripped.

Subject 8
   No major complaints
Subject 9
   A little big, could be smaller. Would use a watch band instead.
Subject 10
   I felt that the wiring was limiting, i.e. my range of motion was perceived as limited. The keypad could easily be inadvertently activated in the sometimes tight workspaces.
Subject 11
   Having to take time to look down at keypad to select next function.
Subject 12
   You had to focus away from what you were doing to enter commands.
Subject 13
   Eyepiece needs to go opposite dominant eye. The computer is a little uncomfortable and the cord on the keypad needs to be longer for easier mobility
Subject 14
   keys vulnerable to being pushed without and control of user. Bulky size
Subject 15
   Inability to back up to last page of instruction
Subject 16
   no response

Appendix I. Responses to Voice Questions

Responses to Voice Questions

Subject Number	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	6	7	5	6	8	8	6	8	3	6
2	7	6	7	7	7	6	7	8	7	6
3	7	9	9	9	6	8	1	7	8	8
4	5	7	7	6	7	8	7	6	8	9
5	9	9	9	9	8	7	5	8	9	9
6	7	7	7	9	7	7	5	7	7	7
7	7	9	9	8	8	8		8	7	8
8	3	4	7	8	5	7	7	8	8	8
9	9	9	9	9	8	8	8	8	8	8
10	7	5	8	4	7	7	6	8	9	9
11	8	7	6	6	8	8	7	9	9	7
12	7	8	8	8	8	7	7	7	8	9
13	8	7	8	6	8	8	8	9	9	9
14	8	7	7	6	8	6	5	4	7	9
15										
16	7	6	8	9	7	7		4	9	5
Mean	7.00	7.13	7.60	7.33	7.33	7.33	6.08	7.27	7.73	7.80
Standard Deviation	1.46	1.45	1.14	1.53	0.87	0.70	1.77	1.48	1.48	1.28

Open-ended Questions

Question 1 - What did you like about voice input?

Subject 1
> easy to use

Subject 2
> Didn't need to push buttons or carry TO's; this kept hands open to do job more efficiently

Subject 3
> very easy to use. Could work on task and look at info

Subject 4
> it was fast

Subject 5
> No keypad or associated wires providing free arm motions and free use of both hands at all times

Subject 6

You could stay focused on what you were viewing while giving commands to move forward in the T.O.

Subject 7

Freed up hands for use

Subject 8

Don't have to flip through a Job Guide

Subject 9

Much better than keypad, for 2 reasons: 1) Increased speed performing tasks, 2) Hands free, which contributes directly to one.

Subject 10

The hands off aspect. When ohm checking a wire I could step without losing grip.

Subject 11

No distraction of having to push buttons to get responses

Subject 12

It made it simple to continue work while having both hands free for other tasks.

Subject 13

liked it very well

Subject 14

It allows full use of hands. This will enable technician to use hands and view tech data at same time. Also, wind would not blow pages in T.O.

Subject 15

no response

Subject 16

Easy, could keep your hands free to perform the necessary tasks, hold tools, and complete job

Question 2 - What did you not like about voice input?

Subject 1

could not talk while you work

Subject 2

Outside noise interference caused computer to "stall out"

Subject 3

headset too large

Subject 4

it responded to one command that wasn't mine (it responded to outside noise)

Subject 5

no response

Subject 6
        Time in programming or recording your voice and occasionally your commands weren't recognized, but that really wasn't much of a problem.
Subject 7
        Its sensitivity
Subject 8
        Too many steps, when completed with a screen it should automatically step to the next screen.
Subject 9
        Nothing.
Subject 10
        I think aloud. I had to curb my desire to speak and do so only when needed to by the system.
Subject 11
        Good system
Subject 12
        Difference in background noise sometimes affected the voice recognition.
Subject 13
        I feel that if you had a small menu of the commands that it would make it a lot easier for people to use, or if you would say the work (commands) it would give you a menu of your programmed commands.
Subject 14
        Had problems with computer recognizing voice commands. Ambient noise was created near end of task and this caused some problems.
Subject 15
        no response
Subject 16
        commands didn't always match the required screen inputs to continue the tasks or switch pages. This made it (completing the task) slower.

Appendix J. Responses to Overall System Questions

Responses to final Questions

Question 1 - What did you like about the image in the eye-piece?

Subject 1
      did not care for the eyepiece
Subject 2
      very clear - easy to read
Subject 3
      the idea is good but the eyepiece needs to be improved
Subject 4
      it was a good image
Subject 5
      provided up front instructions while in work. No pages to mess with
especially on a windy day.
Subject 6
      It made viewing the TO's quicker
Subject 7
      That all data was right in front of your face
Subject 8
      Better than holding a T.O.
Subject 9
      no response
Subject 10
      small, hands off.
Subject 11
      Always there to see
Subject 12
      Small enough not to be in the way but large enough to read
Subject 13
      hands free, always looking forward
Subject 14
      It is available for constant viewing. It is not going to lose a page marker in
windy conditions
Subject 15
      no response
Subject 16
      no response

Question 2 - What did you not like about the image on the eye-piece?

Subject 1
    it's hard for me to focus on screen, being farsighted
Subject 2
    eyepiece attachment device shifted with helmet
Subject 3
    color needs to be added. had trouble with focus
Subject 4
    no response
Subject 5
    small, no color, blocked normal vision in one eye.
Subject 6
    it was small and some of the diagrams were a little difficult to see.
Subject 7
    The placement had to be just right in order to read the monitor
Subject 8
    Image was not fully focused
Subject 9
    Too small and fuzzy
Subject 10
    The image would sometimes move out of range, therefore I would
sometimes need to readjust the eyepiece.
Subject 11
    Hard to focus on at times.
Subject 12
    no response
Subject 13
    hard to focus, in your field of view if you're removing small items such as
screws, etc.
Subject 14
    hard to see pins at times with eyepiece in place
Subject 15
    no response
Subject 16
    no response

Question 3 - What did you like about the "suite" of eye-piece, vest, input devices, etc..?

Subject 1
    no response

Subject 2
    vest need to be trimmed down - but other than that, for a "test equipment",
it's ok
Subject 3
    basic vest is good
Subject 4
    I liked the keypad the best
Subject 5
    mobile, no books to carry. All information is contained in one suite.
Subject 6
    It wasn't as cumbersome as I thought it would be.
Subject 7
    The mobility
Subject 8
    Nothing
Subject 9
    The fact you could move freely about the a/c while performing tasks.
Subject 10
    Self contained.
Subject 11
    Headset will protect your head from aircraft when you're trying to focus on
eyepiece. Ow!
Subject 12
    Overall it was satisfactory
Subject 13
    lightweight
Subject 14
    the vest is self contained
Subject 15
    no response
Subject 16
    no response

Question 4 - What did you not like about the "suite" of eye-piece, vest, input devices, etc..?

Subject 1
    not too much, cannot function freely with everything hanging off the vest.
Also wearing the head gear over a period of time becomes too much
Subject 2
    cables needed to be moved so I would not trip over them
Subject 3
    too heavy

J-3

Subject 4
    head gear too big if trying to get into small places
Subject 5
    Too bulky. Restricted freedom of movement
Subject 6
    limited mobility
Subject 7
    Bulkiness
Subject 8
    Too bulky and restrictive, not way it could be worn in wartime environment with flack vest and chem gear.
Subject 9
    Could be less bulky. Could be dual eyepiece headset w/boom mike attached, pack of power source/CPU on hip
Subject 10
    It was bulky, hot. The headpiece did not fit well in some workspaces, the vest would be in the way. I felt that the entire setup was "fragile".
Subject 11
    Bulky. Hard to do tasks that may require you to lay or get into tight places.
Subject 12
    It seemed a little heavy and a little awkward.
Subject 13
    cables in your way when walking, longer cables on keypad
Subject 14
    This is not as mobile as you may have to be in some cases.
Subject 15
    no response
Subject 16
    no response

Question 5 - Which input device would you prefer to use on a daily basis?

Subject 1
    keypad
Subject 2
    battery pack w/voice input
Subject 3
    voice
Subject 4
    keypad
Subject 5
    voice

Subject 6
    keypad
Subject 7
    voice
Subject 8
    Doesn't matter it is used with the lap-top instead of the eyepiece.
Subject 9
    voice
Subject 10
    Keypad, it is faster to hit three keys than it is to say three words, i.e. when the notes come up on the screen, it takes less time to get to what you need.
Subject 11
    Voice interface
Subject 12
    Voice recognition
Subject 13
    need to use a little more to determine
Subject 14
    probably voice command
Subject 15
    no response
Subject 16
    no response

Question 6 - Do you have any suggestions for future study?

Subject 1
    work on reducing the size of the vest and its components, and doing away with the eyepiece. Maybe come up with a screen that fits on the vest
Subject 2
    no response
Subject 3
    use testing like this study. Could be a great tool for guardsmen
Subject 4
    if you go with voice use enclosed mic
Subject 5
    no response
Subject 6
    Something that's easier to view and that would be easier to remove from your line of sight so you can do the job at hand (possibly fly up glasses).
Subject 7
    no response

Subject 8

System with the lap top would be ideal for hard broke aircraft or with discrepancies that are integrated with several systems and is using in a controlled environment such as a hangar or arch, etc.

Subject 9

Smaller, com/headset plugging into hip pack.

Subject 10

The keypad is intuitive, therefore has less learning curve. The voice is new. In combination with initial use of system it is harder to learn both "voice" and "system" than "keypad" and "system".

Subject 11

To make the suite less bulky and be able to crawl all over the aircraft.

Subject 12

Seeing how diagrams and schematics would look.

Subject 13

a better eye piece

Subject 14

schematics

Subject 15

no response

Subject 16

no response

Question 7 - Do you have any suggestions for future hardware improvements?

Subject 1

no response

Subject 2

develop better battery pack (long life) so cables do not have to be used (power supply)

Subject 3

add mouse to keypad. Add the option to 'go back' on software. Add eyecup to the eyepiece. Color could be used to show warnings in task

Subject 4

it would be nice to have memory storage on unit if main frame goes down

Subject 5

no response

Subject 6

Hardware may be made smaller and encased in the jacket so it would be more durable

Subject 7

make a way to align the eyepiece over your eye.

Subject 8

Eyepiece would be ok if the entire "suite" could be contained in the head gear.

Subject 9

no response

Subject 10

Perhaps having the keypad higher on arm. less restriction and the pad won't interfere with work, plus you won't need to twist your wrist to use the pad, i.e. you won't need to let loose of what you may be holding. The eyepiece could be better. When a task involves multiple people, it would be nice to be able to see at what step the other person is on. I.e. when I finish a step and the next person needs to do something, their screen would alert them and vice-versa. Having a laptop link, so when you are "training" the trainer or trainee can follow along. Will the keypad be compatible with gas mask? Perhaps incorporate the mic in a hardshell in order to isolate the "noise". Crew chiefs do so in engine on conditions. Some maintenance tasks require engine on.

Subject 11

Eyepiece improvements

Subject 12

Headpiece more lightweight.

Subject 13

Subject 14

eyepiece needs work.

Subject 15

no response

Subject 16

no response

Appendix K. Human Use Release Form

INFORMATION PROTECTED BY THE PRIVACY ACT OF 1974

CONSENT FORM

AFIT / ARMSTRONG LABORATORY VOICE RECOGNITION STUDY

1. You are invited to participate in a study to help evaluate the user interface for the Integrated Maintenance Information System (IMIS). The IMIS will provide maintenance personnel with one computer system capable of accessing all information they need to do their jobs. This study will evaluate two potential user interfaces for the system. This study will compare user performance with IMIS while using a keypad to manipulate the system, to user performance using a voice recognition capability to manipulate the system.

2. Your participation in this study will require you to wear a vest-mounted computer with a head-mounted display device and complete two maintenance troubleshooting tasks. One task will be performed using a keypad that will be attached to your wrist. The other task will be performed using a microphone activated voice recognition package.

3. Your participation will not involve risks greater than you encounter performing your normal duties.

4. Your participation in this study will help us to ensure that the IMIS is designed to meet your needs. The ultimate benefit of this project will be to make maintenance personnel more effective and make their jobs easier.

5. The only other way to obtain the required information would be to conduct studies in a laboratory setting using non-maintenance personnel. These people would not be representative of maintenance personnel, and the information gathered would not reflect the true needs of maintenance personnel.

6. I, _____, am participating because I want to. The decision to participate in this study is completely voluntary on my part. No one has coerced or intimidated me into participating in this program.

7. _____ has adequately answered any and all questions I have asked about this study, my participation, and the procedures involved, which are set forth above, and which I have read. I understand that the graduate students conducting the study will be available to answer any questions concerning procedures throughout this study. I understand that if significant new findings develop during the course of this research which may relate to my decision to continue participation, I will be

K-1

informed. I further understand that I may withdraw this consent at any time and discontinue further participation in this study without repercussion.

Subject's Signature

_____

8. I understand that my entitlement to medical care or compensation in the event of injury are governed by federal laws and regulations, and that if I desire further information, I may contact one of the program administrators. I understand that I will not be paid for my participation in this study.

9. I understand that my participation in this study may be photographed, filmed, or audio/videotaped. I consent to the use of these media for training purposes and understand that any release of records of my participation in this study may only be disclosed according to federal law, including the Federal Privacy Act, 55 U. S. C. 552a, and its implementing regulations. This means that personal information will not be disclosed to an unauthorized source without my permission.

10. I FULLY UNDERSTAND THAT I AM MAKING A DECISION WHETHER OR NOT TO PARTICIPATE. MY SIGNATURE INDICATES THAT I HAVE DECIDED TO PARTICIPATE HAVING READ THE INFORMATION PROVIDED ABOVE.

_____   _____
VOLUNTEER SIGNATURE AND SSN   DATE

_____   _____
PRINCIPAL INVESTIGATOR SIGNATURE   DATE

_____   _____
WITNESS SIGNATURE   DATE

INFORMATION PROTECTED BY THE PRIVACY ACT OF 1974

Authority: 10 U. S. C. 8012, Secretary of the Air Force; powers and duties delegation by; implemented by DOI 12-1, Office Locator.

Purpose is to request consent for participation in an approved maintenance research study. Disclosure is voluntary.

Routine Use: Information may be disclosed for any of the blanket routine uses published by the Air Force and reprinted in AFP 12-36 and in Federal Register 52 FR 16431.

K-2

References

Allport, D. A., et. al. "On the Division of Attention: A Disproof of the Single Channel Hypothesis," Quarterly Journal of Experimental Psychology, 24: 225-235 (1972).

Berardinis, Lawrence A. "Look What's Talking," Machine Design, 65: 46-51 (27 August 1993).

Carney Lisa A., and Roger A. Quinto. Evaluation of Access/Navigational Features of a Graphical-User Interface Installed on a Portable Maintenance Aid. MS Thesis, AFIT/GLM/LAL/93S-9. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1993 (AD-A275 791).

Clark, Philip W., et. al. How CALS Can Improve the DoD Weapon System Acquisition Process. Logistic Management Institute, Bethesda MD, November 1992 (AD-A259 928).

Cochran, D. J., et. al. "An Evaluation of the Strengths, Weaknesses, and Uses of Voice Input Devices," Proceedings of the Human Factors Society-24th Annual Meeting. Los Angeles CA, 1980.

Conover, W. J. Practical Nonparametric Statistics (Second Edition). New York: John Wiley and Sons, 1980.

Davenport, Gary. VoiceAssist User's Guide. Creative Technologies Ltd., 1993.

Department of the Air Force. Air Force CALS Business Practices Technical Summary. Contract F33657-92-D-2171. Dayton OH: RJO Enterprises, June 1993 (AD-B174 488).

Elster, Richard S. The Effects of Certain Background Noises on the Performance of a Voice Recognition System. Naval Postgraduate School, Monterey CA, September 1980 (AD A106138).

Everett, Stephanie S., et. al. Talking to InterFIS: Adding Speech Input to a Natural Language Interface. Navy Center for Applied Research in Artificial Intelligence, Information Technology Division, Washington DC, September 1992 (AD-A255865).

Friend, Jeffrey A. and Randy S. Grinstead. <u>Comparative Evaluation of a Monocular Head Mounted Display Device Versus a Flat Screen Display Device in Presenting Aircraft Maintenance Technical Data</u>. MS thesis, AFIT/GSM/LSM/92S-8. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1992 (AD-A258 684).

Karl, Lewis R., et. al. "Speech Versus Mouse Commands for Word Processing: An Empirical Evaluation," <u>International Journal of Man-Machine Studies, 39</u>: 667-687 (October 1993).

Larochelle, S. "Some Aspects of Skilled Movements in Typewriting," in H. Bourma and D.G. Bouwhuis, Eds. <u>Attention and Performance X: Control of Language Processing</u>. Hillsdale NJ: Erlbaum, 1984.

Leggett, J. and G. Williams. "An Empirical Investigation of Voice as an Input Modality for Computer Programming," <u>International Journal of Man-Machine Studies, 21</u>: 493-520 (September 1984).

Manaris, Bill Z., et. al. "Developing a Natural Language Interface for the UNIX Operating System," <u>SIGCHI Bulletin, 26</u>: 34-39 (April 1994).

Martin, Gale L. "The Utility of Speech Input in User-Computer Interfaces." <u>International Journal of Man-Machine Studies, 30</u>: 355-375 (April 1989).

Masquelier, Barbara L. <u>Comparative Evaluation of Monocular Display Devices Relative to Portable Computers for Display of Aircraft Maintenance Technical Data</u>. MS thesis, AFIT/GSM/LSM/91S-18. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1991 (AD-B161 858).

<u>Microsoft Excel User's Guide</u>. Version 5.0. Microsoft Corporation, 1994.

Neter, John and others. <u>Applied Linear Statistical Models</u> (Second Edition). Homewood IL: Irwin, Inc., 1985.

Nielson, Jakob, and Jonathan Levy. "Measuring Usability: Preference vs. Performance." <u>Communications of the ACM, 37</u>: 66-75 (April 1994).

Nye, J. M. "Human Factors Analysis of Speech Recognition Systems," <u>Speech Technology, I</u>: 50-57 (1982).

Pausch, R. and J. H. Leatherby. "A Study Comparing Mouse Only Input versus Mouse Plus Voice Input for a Graphical Editor," <u>Journal of American Voice Input/Output Society, 9</u>: 2-9 (1991).

Poock, Gary K. Experiments With Voice Input for Command and Control: Using Voice Input to Operate a Distributed Computer Network. Naval Postgraduate School, Monterey CA. April 1980 (AD A091055).

Poock, Gary K. and B. Jay Martin. Effects of Emotional Stress and Perceptual Motor Stress on a Voice Recognition System's Accuracy: An Applied Investigation. Naval Postgraduate School, Monterey CA. February 1984 (AD-A142 554).

Poock, Gary K. and E. F. Roland. Voice Recognition Accuracy: What is Acceptable? Naval Postgraduate School, Monterey CA. November 1982 (AD-A129 930).

Quill, Lauri. "Current Trends and Developments in Human-Computer Interface Technology," White Paper. Armstrong Laboratory, Wright-Patterson AFB OH. 31 August 1993.

Reed, Lockwood W. Voice Interactive Systems Technology Avionics (VISTA) Program. Army Avionics Research and Development Activity, Fort Monmouth NJ. 18 June 1982 (AD-A117 288).

Schmandt, C., et. al. "Augmenting a Window System With Speech Input," IEEE Computer, 23: 50-56 (August 1990).

Statistix User's Manual. Version 4.0. Analytical Software, 1992.

Sternberg, S., et al. "The Latency and Duration of Rapid Movement sequences: Comparisons of Speech and Typewriting," in G. E. Stelmach, Ed. Information Processing in Motor Control and Learning. London: Academic Press, 1978.

Thomas, Donald L. Integrated Maintenance Information System: User Field Demonstration and Test Executive Summary. Air Force Human Resources Laboratory, Logistics and Human Factors Division, Wright-Patterson AFB OH. March 1995 (AL/HR-TR-1995-0033).

Thomas, Donald L. and Jeffrey D. Clay, 1st Lt, USAF. Computer-Based Maintenance Aids for Technicians: Project Final Report. Air Force Human Resources Laboratory, Logistics and Human Factors Division, Wright-Patterson AFB OH. August 1988 (AFHRL-TR-87-44).

Triesman, A. and A. Davies. "Divided Attention to Ear and Eye," in S. Kornblum, Ed. Attention and Performance IV. New York: Erlbaum. 1973.

Visick, D., et. al. "The Use of Simple Speech Recognizers in Industrial Applications," Proceedings of INTERACT '84, First IFIP Conference on Human-Computer Interaction. London England. 1984.

Wickens, C. D. "The Structure of Attentional Resources," in R. Nickerson and R. Pew, Eds. Attention and Performance VIII. New York: Erlbaum. 1980.

Wickens C. D., et al. "Multiple Resources, Task-Hemispheric Integrity, and Individual Differences in Time Sharing," Human Factors, 23: 211-230 (1981).

Chapman Vita

David A. Chapman was born on 2 August 1963 in Springfield, Ohio. He graduated from Ohio University in 1986 with a Bachelor of Science Degree in Mechanical Engineering. David entered Officer Training School in September 1986 and received his commission on 12 December 1986. After completing the Aircraft Maintenance Officer Course (AMOC) at Chanute AFB in May 1987, he was assigned to the 347th Tactical Fighter Wing at Moody AFB, Georgia. He held a variety of positions within the Deputy Commander for Maintenance (DCM) complex, including Assistant Officer in Charge (OIC) of the 68th Aircraft Maintenance Unit (AMU), Aircraft Generation Squadron Assistant Maintenance Supervisor, OIC of the 69th AMU, and Maintenance Supervisor for Moody's 1989 Gunsmoke Team.

In March 1992, he was assigned to Air Forces Iceland, Keflavik Naval Air Station, Iceland with the 57th Fighter Squadron (FS). During his two year tour, he held positions of OIC of the 57th AMU, Assistant Chief of Maintenance for the 57th FS, and Maintenance Supervisor for the 35th Maintenance Squadron. In 1994, he was selected to attend the Air Force Institute of Technology at Wright-Patterson AFB, Ohio. He graduated in 1995 with a Masters degree in Logistics Management.

<div style="text-align:center">

Permanent Address: 422 Reames Ave.
Springfield, OH 45505

</div>

Simmons Vita

Captain James R. Simmons is from Houston, Texas. He graduated from Clear Lake High School in 1987. He earned his Bachelor of Science degree in Engineering from the United States Air Force Academy in 1991. After receiving his commission into the United States Air Force and completing the Aircraft Maintenance Officers Course (AMOC), Captain Simmons was assigned to the 347th Fighter Wing at Moody AFB, Georgia.

During his tour at Moody AFB, Captain Simmons filled a variety of maintenance positions in support of the F-16C/D aircraft. These positions included Maintenance Supervisor in the 347th Component Repair Squadron, Assistant Squadron Maintenance Officer in the 69th and 307th Fighter Squadrons, Squadron Section Commander in the 347th Maintenance Squadron (MS), and Officer in Charge of both the Armament Systems Flight and the Aerospace Ground Equipment Flight in the 347th MS. In 1994, Captain Simmons entered the Air Force Institute of Technology at Wright-Patterson AFB, Ohio, and graduated in 1995 with a Masters degree in Logistics Management. He was subsequently assigned to the C-17 System Program Office as an Integrated Logistics Support Manager.

Permanent Address: 72 Wells Dr.
Dayton, OH 45431

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE September 1995	3. REPORT TYPE AND DATES COVERED Master's Thesis	

4. TITLE AND SUBTITLE
A COMPARATIVE EVALUATION OF VOICE VERSUS KEYPAD INPUT FOR MANIPULATING ELECTRONIC TECHNICAL DATA FOR FLIGHT LINE MAINTENANCE TECHNICIANS

5. FUNDING NUMBERS

6. AUTHOR(S)

David A. Chapman, Captain, USAF
James R. Simmons, Captain, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology,
WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GAL/LAL/95S-2

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AL/HRGO
WPAFB OH 45433-7604

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Interactive Electronic Technical Manuals will soon become a requirement for aircraft maintenance technicians. An important aspect in their development is the selection of an input device that will enhance, rather than impede, technician performance. The purpose of this thesis was to evaluate two types of input devices that can be used: a voice recognition input and a keypad input. Studies to date have evaluated the superiority of digital data over paper data, and advantages of using a Head Mounted Display Device over a flat screen laptop computer. No research has evaluated the input device. An experiment was conducted to determine which interface allowed the technicians to work faster. Sixteen F-16 avionics maintenance technicians from the 178th Tactical Fighter Group, Ohio Air National Guard, performed two parallel tasks using each input device. One task was performed using a keypad input device and another task was performed using a voice recognition input device. Raw data showed no statistical difference in task completion times between input devices. However, when computer processing time was subtracted from the voice task times, there was a slight time difference found. Most importantly, results indicate that the technicians liked the advantages of the voice recognition input device over the keypad input device. The primary conclusion is that voice recognition may be a desirable input configuration and further study is warranted in more stringent environmental conditions.

14. SUBJECT TERMS Aircraft Maintenance, Portable Maintenance Aids, Electronic Technical Manuals HMDD, Voice Recognition, Human-Computer Interface	15. NUMBER OF PAGES 105
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL